Attorney Docket No.: 016976-000810US

PATENT APPLICATION

SURFACE EXPRESSION OF BIOLOGICALLY ACTIVE PROTEINS IN BACTERIA

Inventor(s): Chia-Hwa Chang, citizen of Taiwan, residing at 231 Easy St. #5, Mountain View, CA 94043

Xiaowen Liu, citizen of China, residing at 10164 Imperial Ave., Cupertino, CA 95014

John A. Lewicki, citizen of USA, residing at 308 Escobar Ave., Los Gatos, CA 95032

Qiang Xu, citizen of USA, residing at 21682 Olive Ave., Cupertino, CA 95014

Assignee: OSEL, INC.
1800 Wyatt Drive, Suite 14
Santa Clara, CA, 95054

Entity: Small

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

SURFACE EXPRESSION OF BIOLOGICALLY ACTIVE PROTEINS IN BACTERIA

CROSS REFERENCE TO RELATED APPLICATIONS

[01] The present application claims benefit of priority to U.S. Provisional Patent Application No. 60/443,619, filed on January 29, 2003, which is incorporated by reference in its entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[02] This invention was made with Government support under Grant No. 2 R44 AI46203-02, awarded by the National Institutes of Health. The government has certain rights in this invention.

BACKGROUND OF THE INVENTION

[03] Surface expression of proteins via covalent linkage with peptidoglycans in Gram-positive bacteria involves unique sorting signals and Sortase-dependent machinery (Mazmanian *et al.*, Science 285:760-763 (1999)). One of the best-studied systems is the *emm6* gene of *Streptococcus pyogenes* that encodes the M6 structural protein (Fischetti *et al.*, 1990. *Mol. Microbiol.* 4:1603-1605 (1990)). The M6 proteins have a signature cell wall sorting signal, the Leu-Pro-X-Thr-Gly (LPXTG) motif, followed by a stretch of hydrophobic amino acids and finally a sequence containing charged residues (KRKEEN), which serves as a cell surface retention signal. These cell wall sorting motifs have been identified in other Gram-positive bacteria including *Staphlyococcus*, *Enterococcus*, and *Listeria*, and *Lactobacillus* (Navarre and Schneewind, *Microbio. Mol. Biol. Rev.* 63:174-229 (1999)), but not in *Lactobacillus* species that colonize the human vagina.

[04] The mucosal membranes of all humans are naturally colonized by bacteria (Tannock. *Clin. Rev. Allergy Immunol.* 22: 231-53 (2002)). Recent scientific evidence has documented the fact that these bacteria interact closely with cells and tissues of the body to regulate natural biological processes. It has become increasingly evident that this mucosal microflora also contributes substantially to numerous diseases affecting cells and tissues of humans.

[05] Generally, domination of the microflora within the vagina and gastrointestinal tract, by lactobacilli and related bacteria, is associated with good health (Redondo-Lopez *et al.*, *Rev. Infect. Dis.* 12: 856-72 (1990); Tannock. *Clin. Rev. Allergy Immunol.* 22: 231-53 (2002)). Natural strains of lactobacilli have been administered for many years as "probiotics" for the purpose of maintaining a healthy microflora within these locations and preventing infection. It is well established that these "healthy bacteria" compete with pathogenic organisms, such as bacteria, viruses and fungi to limit the development and progression of pathogen associated diseases. Nevertheless, this microflora is a fragile and dynamic environment with the natural turnover and disruption of the healthy microflora being associated with the establishment of opportunistic infections. Consequently, approaches to maintain, or even enhance, the integrity and natural properties of the microflora, as a means of preventing or treating disease, would be coveted by the biomedical community.

[06] The mucosal microflora contributes to many local diseases affecting mucosal surfaces. For instance, HIV and other sexually transmitted pathogens must bypass the vaginal mucosa. In addition, the etiology of inflammatory bowel diseases, including ulcerative colitis and Crohn's disease may arise from inappropriate interactions between a disrupted mucosal microflora and cells and tissues of the host. A means of modulating the properties of bacteria within the mucosal flora could aid in the prevention or treatment of these diseases, as well as related conditions affecting mucosal surfaces. Targeting biologically active proteins to the cell wall of these and other organisms could help to treat such diseases.

[07] The present invention addresses these and other problems.

BRIEF SUMMARY OF THE INVENTION

[08] The present invention provides *Lactobacillus* bacteria comprising an expression cassette, the expression cassette comprising a promoter operably linked to polynucleotide encoding a signal sequence and a biologically-active polypeptide, wherein the biologically active polypeptide is linked to a heterologous carboxyl terminal cell wall targeting region and wherein the heterologous carboxyl terminal cell wall targeting region comprises in the following order: a cell wall associated sequence; LPQ(S/A/T)(G/A); and a hydrophobic sequence.

[09] In some embodiments, the cell wall associated sequence comprises at least 50 amino acids. In some embodiments, the cell wall associated sequence comprises at

least 200 amino acids. In some embodiments, the heterologous carboxyl terminal cell wall targeting region further comprises a charged sequence at the carboxyl terminus of region.

[10] In some embodiments, the *Lactobacillus* bacterium is a vagina-colonizing strain. In some embodiments, the bacterium is selected from the group consisting of *L. jensenii*, *L. gasseri*, *L. casei* and *L. crispatus*.

[11] In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQSG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQAG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTA. In some embodiments, the cell wall targeting region comprises SEQ ID NO:7. In some embodiments, the cell wall targeting region comprises SEQ ID NO:8.

[12] In some embodiments, the biologically active polypeptide is expressed in the cell wall of the bacterium. In some embodiments, the biologically-active polypeptide is between 10 and 600 amino acids. In some embodiments, the biologically active protein binds to a pathogen when the biologically active protein is contacted with the pathogen.

[13] In some embodiments, the pathogen is a bacterial pathogen. In some embodiments, the pathogen is a fungal pathogen. In some embodiments, the pathogen is a viral pathogen.

[14] In some embodiments, the viral pathogen is a human immunodeficiency virus (HIV). In some embodiments, the biologically active protein is CD4 or an HIV-binding fragment of CD4. In some embodiments, the biologically active protein is 2D-CD4. In some embodiments, the biologically active protein is cyanovirin-N (CV-N) or a virus-binding fragment of CV-N. In some embodiments, the viral pathogen is herpes simplex virus. In some embodiments, the biologically active protein is herpes simplex virus entry mediator C (HveC) or a virus-binding fragment of HveC.

[15] In some embodiments, the biologically active polypeptide is released from the *Lactobacillus* bacterium. In some embodiments, the biologically active polypeptide is anchored to the cell wall of the *Lactobacillus* bacterium.

[16] The present invention also provides methods of expressing a biologically active polypeptide in the cell wall of a *Lactobacillus* bacterium. In some embodiments, the method comprises providing a *Lactobacillus* bacterium comprising an expression cassette, the expression cassette comprising a promoter operably linked to a polynucleotide encoding a signal sequence and a biologically-active polypeptide, wherein the

biologically active polypeptide is linked to a heterologous carboxyl terminal cell wall targeting region and wherein the heterologous carboxyl terminal cell wall targeting region comprises in the following order: a cell wall associated sequence; LPQ(S/A/T)(G/A); and a hydrophobic sequence; and culturing the bacterium under conditions to induce expression of the polypeptide, thereby expressing a biologically active polypeptide in the cell wall of the *Lactobacillus* bacterium.

[17] In some embodiments, the cell wall associated sequence comprises at least 50 amino acids. In some embodiments, the cell wall associated sequence comprises at least 200 amino acids.

[18] In some embodiments, the heterologous carboxyl terminal cell wall targeting region further comprises a charged sequence at the carboxyl terminus of region. In some embodiments, the providing step comprises transferring the expression cassette into the bacterium.

[19] In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQSG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQAG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTA. In some embodiments, the cell wall targeting region comprises SEQ ID NO:7. In some embodiments, the cell wall targeting region comprises SEQ ID NO:8.

[20] In some embodiments, the cell wall targeting region comprises at least 200 amino acids.

[21] In some embodiments, the bacterium is vagina-colonizing strain. In some embodiments, the bacterium is selected from the group consisting of *L. jensenii*, *L. gasseri*, *L. casei,* and *L. crispatus*. In some embodiments, the biologically-active polypeptide is between 10 and 600 amino acids. In some embodiments, the biologically active protein binds to a pathogen when the biologically active protein is contacted with the pathogen.

[22] In some embodiments, the pathogen is a bacterial pathogen. In some embodiments, the pathogen is a fungal pathogen. In some embodiments, the pathogen is a viral pathogen.

[23] In some embodiments, the viral pathogen is HIV. In some embodiments, the biologically active protein is CD4 or an HIV-binding fragment of CD4. In some embodiments, the biologically active protein is 2D-CD4. In some embodiments, the biologically active protein is cyanovirin-N or a virus-binding fragment of cyanovirin-N. In

some embodiments, the biologically active protein is herpes simplex virus entry mediator C (HveC) or a virus-binding fragment of HveC.

[24] In some embodiments, the biologically active polypeptide is released from the *Lactobacillus* bacterium. In some embodiments, the biologically active polypeptide is anchored in the cell wall of the *Lactobacillus* bacterium.

[25] The present invention also provides methods of providing a biologically active protein to a mammalian mucosal surface. In some embodiments, the methods comprise contacting a mucosal surface with a *Lactobacillus* bacterium recombinantly altered to express a signal sequence linked to a biologically-active polypeptide linked to a heterologous carboxyl terminal cell wall targeting region, the heterologous carboxyl terminal cell wall targeting region comprising in the following order: a cell wall associated sequence; LPQ(S/A/T)(G/A); and a hydrophobic sequence, wherein the biologically active polypeptide is expressed in an amount able to be detected in a sample collected from the mucosal surface.

[26] In some embodiments, the cell wall associated sequence comprises at least 50 amino acids. In some embodiments, the cell wall associated sequence comprises at least 200 amino acids. In some embodiments, the heterologous carboxyl terminal cell wall targeting region further comprises a charged sequence at the carboxyl terminus of region. In some embodiments, the Lactobacillus bacterium is selected from the group consisting of *L. jensenii*, *L. gasseri*, *L. casei* and *L. crispatus*.

[27] In some embodiments, the mucosal surface resides within the vagina. In some embodiments, the mucosal surface resides within the gastrointestinal tract.

[28] In some embodiments, the contacting step comprises orally administering the Lactobacillus bacteria. In some embodiments, the contacting step comprises vaginally administering the Lactobacillus bacteria. In some embodiments, the contacting step comprises rectally administering the Lactobacillus bacteria.

[29] The present invention provides expression cassettes comprising a promoter operably linked to polynucleotide encoding a signal sequence and a biologically-active polypeptide, wherein the biologically active polypeptide is linked to a heterologous carboxyl terminal cell wall targeting region, the heterologous carboxyl terminal cell wall targeting region comprising in the following order: a cell wall associated sequence; LPQ(S/A/T)(G/A); and a hydrophobic sequence. In some embodiments, the cell wall associated sequence comprises at least 50 amino acids. In some embodiments, the cell wall associated sequence comprises at least 200 amino acids.

[30] In some embodiments, the heterologous carboxyl terminal cell wall targeting region further comprises a charged sequence at the carboxyl terminus of region.

[31] In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQSG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQAG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTG. In some embodiments, the cell wall targeting region comprises the amino acid sequence LPQTA. In some embodiments, the cell wall targeting region comprises SEQ ID NO:7. In some embodiments, the cell wall targeting region comprises SEQ ID NO:8. In some embodiments, the biologically-active polypeptide is between 10 and 600 amino acids.

[32] In some embodiments, the biologically active protein binds to a pathogen when the biologically active protein is contacted with the pathogen. In some embodiments, the pathogen is a bacterial pathogen. In some embodiments, the pathogen is a fungal pathogen. In some embodiments, the pathogen is a viral pathogen. In some embodiments, the viral pathogen is HIV.

[33] In some embodiments, the biologically active protein is CD4 or an HIV-binding fragment of CD4. In some embodiments, the biologically active protein is 2D-CD4. In some embodiments, the biologically active protein is cyanovirin-N or a virus-binding fragment of cyanovirin-N. In some embodiments, the biologically active protein is herpes simplex virus entry mediator C (HveC) or a virus-binding fragment of HveC. In some embodiments, the cell wall targeting region functions in Lactobacillus.

[34] The present invention also provides vectors comprising an expression cassette comprising a promoter operably linked to polynucleotide encoding a biologically-active polypeptide linked to a heterologous carboxyl terminal cell wall targeting region, the heterologous carboxyl terminal cell wall targeting region comprising in the following order: a cell wall associated sequence; LPQ(S/A/T)(G/A); and a hydrophobic sequence.

DEFINITIONS

[35] A "biologically active protein" refers to an amino acid sequence that has the biological activity (*i.e.*, can participate in the molecular mechanisms) of the amino acid sequence within, or outside of, a native cell. Activity of a protein includes, *e.g.*, its immunogenicity, catalytic activity, binding affinity, etc. Polypeptide vaccines are encompassed by the term "biologically active proteins." Typically, the amino acid sequence

forms the three-dimensional structure formed by the amino acid sequence within or outside of the native cell.

[36] "2D CD4" refers to the first approximately 183 amino acids of human CD4 (Arthos *et al.*, *Cell.* 1989. 57: 469-81 (1989)). CD4 is a cell-surface glycoprotein found on the mature helper T cells and immature thymocytes, as well as monocytes and macrophages. 2D-CD4 binds to HIV-1 gp120 with the same affinity as the intact protein, and contains the binding site for gp120. CD4 contains an amino-terminal extracellular domain (amino acid residues 1 to 371), a transmembrane region (372 to 395) and a cytoplasmic tail (396 to 433).

[37] "Antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

[38] An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

[39] Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab')2, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab')2 may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the F(ab')2 dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (*see*, Paul (Ed.) *Fundamental Immunology*, Third Edition, Raven Press, NY (1993)). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments

either produced by the modification of whole antibodies or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv).

[40] The term "isolated," when applied to a nucleic acid or protein, denotes that the nucleic acid or protein is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in a preparation is substantially purified. In particular, an isolated gene is separated from open reading frames that flank the gene and encode a protein other than the gene of interest. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

[41] The term "nucleic acid" or "polynucleotide" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term "nucleic acid" is used interchangeably with "polynucleotide."

[42] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. As used herein, the terms encompass amino acid chains of any length, including full-length proteins (*i.e.*, antigens), wherein the amino acid residues are linked by covalent peptide bonds.

[43] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. "Amino acid mimetics" refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but which functions in a manner similar to a naturally occurring amino acid.

[44] Amino acids may be referred to herein by either the commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

Two nucleic acid sequences or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence over a comparison window, as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. When percentage of sequence identity is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions, where amino acids residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an

identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated according to, *e.g.*, the algorithm of Meyers & Miller, *Computer Applic. Biol. Sci.* 4:11-17 (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

[45] The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to a sequence or subsequence that has at least 70% sequence identity with a reference sequence. Alternatively, percent identity can be any integer from 40% to 100%. More preferred embodiments include at least: 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98 or 99% compared to a reference sequence (*e.g.*, SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, or SEQ ID NO:8 or fragments thereof) using the programs described herein, such as BLAST using standard parameters, as described below.

[46] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

[47] A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection.

[48] An example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al., J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

[49] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[50] The term "recombinant" or "recombinantly altered" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all.

[51] The term "heterologous" when used with reference to portions of a nucleic acid or a polypeptide indicates that the nucleic acid or polypeptide comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

[52] An "expression cassette" is a nucleic acid, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression cassette can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

BRIEF DESCRIPTION OF THE DRAWINGS

[53] Figure 1 illustrates the structures of three cell wall anchored proteins identified after genomic sequencing of *L. jensenii* 1153. All of the three proteins have LPQTG sorting signal preceding a hydrophobic region and a charged C-terminal tail and possess unique long repetitive sequences. CWA represents putative cell wall associated regions upstream of the LPQTG motif.

[54] Figure 2A-C illustrates cell wall anchor sequences (C14, C191, and C370) resulting from genomic sequencing of *L. jensenii* 1153. The CWA200 region along with anchor motif is underlined. CWA200 represents putative cell wall associated or spanning regions of about 200 amino acids upstream of the LPQTG motif.

[55] Figure 3 illustrates results from western analysis of SDS extractable proteins and cell wall enriched fractions following mutanolysin digestion of transformed *L. jensenii* 1153 when cultured in MRS broth (A) or Rogosa SL broth (B) at 37°C and 5% CO_2. After separation in reducing SDS-PAGE, the proteins were electroblotted to PVDF membranes for probing with monoclonal antibody (mAb) against c-Myc.

[56] Figure 4 illustrates results from western analysis of cell wall enriched fractions following mutanolysin digestion of transformed *L. jensenii* 1153 when cultured in Rogosa SL broth at 37°C and 5% CO_2. After separation in reducing SDS-PAGE, the proteins were electroblotted to PVDF membrane for probing with polyclonal antibodies (pAb) against CD4 (T4-4). The expression constructs contained the following elements: P_{23} promoter-

CbsA signal sequence (CbsAss)-2D CD4 in pOSEL651; P_{23} promoter-CbsAss-2D CD4-CWA200-anchor of C14 sequence in p237; P_{23} promoter-CbsAss-2D CD4-CWA200-anchor of C191 sequence in pOSEL242; P_{23} promoter-CbsAss-2D CD4-CWA200-anchor of C370 sequence in pOSEL249. CWA200 represents approximately 200 amino acids upstream of C-terminal anchor domain.

[57] Figure 5 illustrates results from flow cytometric analysis of *L. jensenii* 1153 harboring plasmids designed for secretion or surface anchoring of 2D CD4. The bacterial cells were probed with rabbit pAb against CD4 (T4-4), and then FITC-conjugated anti-rabbit antibodies (A). Alternatively, the bacterial cells were probed with mAb Sim.4, and then PE-conjugated anti-mouse IgG (B). Controls consisted of unstained cells or cells probed with fluorochrome-conjugated secondary antibodies. The fluorescence density as a measure of antibody binding to bacterial surface was calculated using FLOWJO software.

[58] Figure 6 illustrates that the C-terminal anchor motif of 36-amino acid in length is insufficient to drive surface expression of 2D CD4. (A). Constructs designed for surface expression of 2D CD4 using native anchor sequences in *L. jensenii*. (B). Flow cytometric analysis of *L. jensenii* 1153 harboring pOSEL238 or pOSEL237. The bacterial cells were probed with mAb Sim.4 against CD4, and then phycoerythrin (PE)-conjugated anti-mouse antibodies. Controls consisted of unstained cells or cells probed with PE-conjugated secondary antibodies.

[59] Figure 7 illustrates the surface expression of 2D CD4 in *L. jensenii* 1153 as affected by different number of the repetitive cell wall spanning sequence upstream of the LPQTG sorting signal in C370 sequence. Surface exposed 2D CD4 molecules that adopt a correctly folded conformation were probed with mAb Sim.4 for flow cytometric analysis in the bacterial cells harboring the following plasmid: 175, a negative control; 249, two and a half repeats; 262, no repeat; 268, one repeat; 278, two repeats; 280, four repeats; 281, seven repeats; 276, eight repeats.

[60] Figure 8 illustrates the surface display of c-Myc tagged proteins in a variety of lactobacillus species of human origin. (A). Schematic of pOSEL241 designed for expression of c-Myc tagged CWA200 of C370 sequence under control of P23 promoter and CbsA signal sequence (CbsAss). (B). Western analysis of cell wall enriched fractions following mutanolysin digestion of transformed *L. jensenii*, *L. gasseri*, and *L. casei*. After separation in reducing SDS-PAGE, the proteins were electroblotted to PVDF membrane for probing with mAb against c-Myc. (C). Flow cytometric analysis of human vaginal lactobacillus isolates harboring pOSEL241. The bacterial cells were probed with mAb

against c-Myc, and then phycoerythrin (PE)-conjugated anti-mouse antibodies. Controls consisted of unstained cells or cells probed with PE-conjugated secondary antibodies.

[61] Figure 9 illustrates the effect of point mutations in the LPQTG motif of C14 and C370 sequences on the surface display of 2D-CD4-CWA200 in *L. jensenii* 1153. Bacterial cells were surface-stained by using pre-titered mAb Sim.4 (A) or pAb T4-4 (B), followed by probing with PE-conjugated anti-mouse or FITC conjugated anti-rabbit antibodies. The flow cytometric analysis was performed in a FACScalibur system. The difference between the protein displayed on the cell surface of pOSEL237, pOSEL249, and those in bacterial cells harboring mutagenic constructs was expressed in mean fluorescence intensity. The surface display of 2D CD4 in the bacterial cells harboring pOSEL237 or pOSEL249 was arbitrarily set as 100%.

[62] Figure 10 illustrates schematic diagram of deletion constructs in C-terminal charged tails of C14 and C370 sequences.

[63] Figure 11 illustrates the effect of sequence deletion in the C-terminal charged tails of C14 and C370 on the surface display of 2D CD4-CWA200. Bacterial cells were surface-stained by using pre-titered pAb T4-4 (A) or mAb Sim.4 (B), followed by probing with FITC conjugated anti-rabbit or PE-conjugated anti-mouse antibodies. The binding of antibody to cell wall anchored proteins was analyzed by flow cytometry using a FACScalibur system. The difference between the protein displayed on the cell surface of pOSEL237 or pOSEL249 and those in bacterial cells harboring mutagenic constructs was expressed as mean fluorescence intensity. The surface display of 2D CD4 in the bacterial cells harboring pOSEL237 or pOSEL249 was arbitrarily set as 100%.

[64] Figure 12 illustrates a comparison of activities of secreted 2D CD4-CWA200 in *L. jensenii* 1153 harboring pOSEL237-7 and pOSEL249-10 relative to 2D CD4 from those harboring pOSEL651. CD4 ELISA was designed to recognize proteins that adopt a correct, properly-folded conformation in cell-free conditioned media. Amounts of proteins were normalized based on their immunoreactivity to pAb T4-4. The soluble 2D CD4 proteins released from the bacterial cells harboring pOSEL651 was arbitrarily set as 100%.

DETAILED DESCRIPTION OF THE INVENTION

I. Introduction

[65] The present invention provides novel motifs and methods for expressing heterologous polypeptides on the cell wall of Gram-positive bacteria such as *Lactobacillus*. The motifs of the invention can be fused to a protein of interest and then

expressed as a fusion protein in the bacteria, resulting in targeting, imbedding, and/or surface display of the fusion protein in the cell wall, or releasing the biologically active and stable fusion protein to the extracellular matrix.

[66] The motifs are useful, for instance, for expression of proteins on the cell wall of *Lactobacillus* bacteria that colonize the human mucosa, including the vagina. Exemplary mucosal bacteria include *Lactobacillus* species, such as *L. jensenii*, *L. gasseri*, and *L. casei*.

II. Cell Wall Targeting Regions

[67] To express and target a polypeptide of interest covalently anchored to a cell wall in Gram-positive bacteria such as *Lactobacillus*, the cell wall targeting region is C-terminally linked to a heterologous polypeptide of interest. The cell wall targeting region enabling surface display of heterologous proteins in vaginally-associated lactobacilli as well as other lactobacilli is comprised of four parts: a cell wall associated region, a LPQ(S/A/T)(G/A) sequence, and a hydrophobic sequence, typically in that order. Optionally, the cell wall targeting region will comprise a charged region at or near the carboxyl terminus. The charged region acts as a stop-transfer sequence in the cell membrane, thereby preventing release into the media. Of course, release into the media may still occur if the anchoring sequence is cleaved from the rest of the protein.

A. Cell Wall Associated Region

[68] The cell wall associated region precedes the LPQ(S/A/T)(G/A) sorting signal. The length of the cell wall associated region may vary. The cell wall associated region is typically between 40 and 1,000 amino acids. In some embodiments, the cell wall associated region is at least about 30, 50, 80, 100, 150, 200 or more amino acids. In some embodiments, the cell wall associated region has about 500, 400, 300, 250, 200, 150, 100 or fewer amino acids. In *Lactobacillus jensenii*, a stretch of 95 amino acids containing one tandem repeat in fusion with the C-terminal cell wall sorting signal in pOSEL268 (described in the Examples) enables surface display of CD4. However, approximately 50 amino acids long in M6 protein of *S. pyogenes* was identified based on peptide mapping (Pancholi & Fischetti, *J. Bacteriol.* 170:2618-2624 (1988)), whereas about 90 amino acids of a fibronectin binding protein was postulated in *S. carnosus* (Strauss & Gotz, *Mol. Microbiol.* 21:491-500 (1996)). Thus, sequences about 50 amino acids or less can be functional in *Lactobacillus*.

[69] In some embodiments, the cell wall associated region is hydrophilic. In some embodiments, the cell wall associated region contains imperfect tandem repeats that can vary in length and sequence. For example, the cell wall associated region of *L. jensenii* C370 contains two and a half tandem repeats. However, while tandem repeats may occur in the cell wall associated region, it is not required. For example, the cell wall associated region of C14 contains no repeats. Functionally, the cell wall associated region interacts with and spans the peptidoglycan layer. Accordingly, it is also called a cell wall spanning or attachment domain, acting as a spacer between the protein that is anchored by membrane-associated sortase and the cell wall sorting signal.

[70] The present invention provides cell wall associated regions substantially identical to the C370 sequence KKAEEVKNNSNATQKEVDDATNNLKQAQNDLDGQTTDKSKLDEAIKSADDTKSTD KYNNASDDTKSKFDEALKKAEEVKNNSNATQKEVDDATKNLKQAQNDLDGQTTN KDAINDAIKDANNAKGTDKYNNASDDTKSKFDDALKKAEDVKNDSNANQKEVDD ATKNLKNTLNNLKGQPAKKANLIASKDNAKIHKQTL (SEQ ID NO:4). In some cases, the cell wall associated region comprises at least about 40, 50, 75, 90, 100, 120, 150, 175, 200 amino acid fragments of the C370 sequence. For example, an active cell wall associated fragment can comprise the following sequence: GQTTNKDAINDAIKDANNAKGTDKYNNASDDTKSKFDDALKKAEDVKNDSNANQK EVDDATKNLKNTLNNLKGQPAKKANLIASKDNAKIHKQTL (SEQ ID NO:5). The C370 sequence (SEQ ID NO:4) comprises 75 charged amino acid residues (K, R, D, E) and lacks Pro-Gly rich sequences.

[71] In some embodiments, the cell wall associated regions is substantially identical to the C14 sequence: VTRTINVVDPITGKISTSVQTAKFTREDKNSNAGYTDPVTGKTTMNPWTPAKQGLRA VNVEQIKGYVAKVDGNVDAVVVTPDSANMVVTITYQANKPEGQNITVKKDTVPDP ADGIKNKDDLPDGTKYTWKEVPDVNSVGEKTGIVTVTFPDGTSVDVKVTVYVDPVV ESNRDTLSKEANTGNTNVAKAATVTSSKVESKKT (SEQ ID NO:6). In some cases, the cell wall associated region comprises at least about 40, 50, 75, 90, 100, 120, 150, 175, 200 amino acid fragments of the C370 sequence. SEQ ID NO:6 comprises 51 charged amino acid residues (K, R, D, E).

[72] In some cases, the cell wall associated region is derived from bacteria other than Lactobacillus or from a Lactobacillus strain not associated with the vagina.

B. LPQ(S/A/T)(G/A)

[73] The sequence LPQ(S/A/T)(G/A) acts as a cell wall sorting signal in vaginally associated strains of Lactobacillus. At least one copy of the motif LPQ(S/A/T)(G/A) is in the cell wall targeting region. The parentheses in the motif indicate alternative amino acids in that position (*e.g.*, LPQSG, LPQAG, LPQTG, LPQSA, LPQAA, LPQTA).

C. Hydrophobic sequences

[74] The carboxyl terminus of a polypeptide to be anchored in the cell wall comprises a hydrophobic region that functions to span the bacterial membrane. The hydrophobic region comprises at least about 50%, and in some embodiments, at least 60%, 70%, 80% or 90% hydrophobic amino acids. Naturally occurring hydrophobic amino acids include alanine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan and valine. Some less hydrophobic amino acids, including glycine, threonine, and serine, can also constitute part of these sequences (*see, e.g.*, Pallen *et al.*, *Trends Microbiol.* 9:97-101 (2001)). Hydrophobic sequences generally are between about 10 and about 30 amino acids and sometimes 13 and 24 amino acids in length among available LPXTG-containing substrates for sortase-like proteins (Pallen *et al.*, *Trends Microbiol.* 9:97-101 (2001)). Exemplary hydrophobic sequences include, *e.g.*, V^{1740}GILGLAIATVGSLLGLGV^{1758} in C14 and P^{1877}LTAIGIGLMALGAGIFA^{1894} in C370.

[75] Alternatively, the hydrophobic regions of any cell wall anchored protein from a Gram positive bacterium can be used. Alternate hydrophobic sequences include, *e.g.*, those described in Figure 1 of U.S. Patent No. 5,821,088 or substantially identical sequences. Additional sequences are also depicted in Table 2 of Pallen *et al*, *Trends Microbiol.* 9: 97-100 (2001).

D. Charged sequences

[76] A charge region can be optionally present at the carboxyl terminus of a cell wall targeted protein, typically immediately following the hydrophobic membrane spanning region. The presence of a carboxyl terminal charged region anchors the polypeptide to the membrane, thereby greatly reducing the amount of protein that dissociates from the membrane and escapes into the media. The charged region comprises at least 40%, and in some embodiments, at least 50%, 60%, 70%, 80% or 90%, charged amino acids. Naturally occurring charged amino acids include arginine, histidine, lysine, aspartic acid and glutamic

acid. Charged sequences can be between, *e.g.*, 2 and 20 amino acid residues and in some embodiments are between 4 and 12 or between 5 and 11 amino acids in length. Exemplary charged sequences include, *e.g.*, $K^{969}KRKED^{974}$ in C191, $R^{1760}KKRQK^{1765}$ in C14, and $K^{1895}KKRKDDEA^{1903}$ in C370.

[77] Alternatively, the charged regions of any cell wall anchored protein from a Gram positive bacterium can be used. Alternate charged sequences include, e.g., those described in Figure 1 of U.S. Patent No. 5,821,088 or substantially identical sequences. Additional sequences are also depicted in Table 2 of Pallen *et al*, *Trends Microbiol.* 9: 97-100 (2001).

III. Recombinant Techniques

A. Molecular Biology Methods

[78] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (3rd ed. 2001); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

[79] For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Proteins sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

[80] Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et al.*, *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

[81] The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, *e.g.*, the chain termination method for sequencing double-stranded templates of Wallace *et al.*, *Gene* 16:21-26 (1981).

B. Cloning Methods for the Isolation of Nucleotide Sequences Encoding Desired Proteins

[82] In general, the nucleic acids encoding the subject proteins are cloned from DNA libraries that are made from cDNA or genomic DNA. The particular sequences can be located by hybridizing with an oligonucleotide probe, the sequence of which can be derived from the sequences disclosed herein or are known in the art, which provide a reference for PCR primers and defines suitable regions for isolating gene-specific probes. Alternatively, where the sequence is cloned into an expression library, the expressed recombinant protein can be detected immunologically with antisera or purified antibodies made against a polypeptide of interest, including those disclosed herein.

[83] Methods for making and screening genomic and cDNA libraries are well known to those of skill in the art (*see, e.g.*, Gubler & Hoffman, *Gene* 25:263-269 (1983); Benton & Davis, *Science*, 196:180-182 (1977); and Sambrook, *supra*). Cells expressing a protein of interest are useful sources of RNA for production of a cDNA library.

[84] Briefly, to make the cDNA library, one should choose a source that is rich in mRNA. The mRNA can then be made into cDNA, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. For a genomic library, the DNA is extracted from a suitable tissue or cell and either mechanically sheared or enzymatically digested to yield fragments of preferably about 5-100 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro,* and the recombinant phages are analyzed by plaque hybridization. Colony hybridization is carried out as generally described in Grunstein *et al.*, *Proc. Natl. Acad. Sci. USA.*, 72:3961-3965 (1975).

[85] An alternative method combines the use of synthetic oligonucleotide primers with polymerase extension on an mRNA or DNA template. This polymerase chain reaction (PCR) method amplifies the nucleic acids encoding the protein of interest directly from mRNA, cDNA, genomic libraries or cDNA libraries. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acids encoding specific proteins and express said proteins, to synthesize nucleic acids that will be used as probes for detecting the presence of mRNA encoding a polypeptide of the invention in physiological samples, for nucleic acid sequencing, or for other purposes (*see*, U.S. Patent Nos. 4,683,195 and

4,683,202). Genes amplified by a PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

[86] Appropriate primers and probes for identifying the genes encoding a polypeptide of the invention from tissues or cell samples can be derived from the sequences described in the art. For a general overview of PCR, *see*, Innis *et al. PCR Protocols: A Guide to Methods and Applications*, *Academic Press*, San Diego (1990).

[87] A polynucleotide encoding a polypeptide of the invention can be cloned using intermediate vectors before transformation into *Lactobacillus*. These intermediate vectors are typically prokaryote vectors or shuttle vectors.

C. Transformation Techniques

[88] Appropriate bacterial host strains are selected for, *e.g.* their transformation ability, ability for heterologous protein expression, and/or ability to colonize on mucosal surfaces. The bacterial host will be rendered competent for transformation using standard techniques, such as the rubidium chloride method or electroporation (*see, e.g.*, Wei, *et al.*, *J. Microbiol. Meth.* 21:97-109 (1995).

[89] Transformation of *L. jensenii* by electroporation can be performed by modifying standard methods as described in, *e.g.*, Luchansky *et al.* (*J. Dairy Sci.* 74: 3293-3302 (1991); Chang *et al.*, *Proc. Natl. Acad. Sci. USA*. 100:11672-11677 (2003)). Briefly, freshly inoculated *L. jensenii* are cultured in broth (*e.g.*, to 0.6-0.7 at OD_{600} at 37°C and 5% CO_2). The bacterial cells are harvested, washed and re-suspended in a cold (*e.g.*, 4°C) solution of sucrose and $MgCl_2$. Competent cells are then mixed with DNA and placed in a chilled gap cuvette and electroporated. Afterward, cells are allowed to recover in pre-warmed broth (*e.g.*, for about two hours at 37°C), prior to being plated on selective agar plate containing an antibiotic other selective agent.

D. Expression

[90] Expression cassettes of the invention can include a variety of components to regulate expression and localization of the polypeptides of the invention. For example, expression cassettes can include promoter elements, sequences encoding signal sequences, a coding sequence for the polypeptide of interest and anchor sequences.

[91] Expression of the heterologous polynucleotides or polypeptides can be constitutive (*e.g.*, using P59 (Van der Vossen *et al.*, *Appl. Environ. Microbiol.* 58:3142-3149

(1992)) or P23 (Elliot *et al.*, *Cell* 36:211-219 (1984)) promoters, or *Lactobacillus*-derived native promoters of even higher strength). Alternatively, expression can be under the control of an inducible promoter. For example, the *Bacillus* amylase (Weickert *et al.*, *J. Bacteriol.* 171:3656-3666 (1989)) or xylose (Kim *et al. Gene* 181:71-76 (1996)) promoters as well as the *Lactococcus* nisin promoter (Eichenbaum *et al*, *Appl. Environ. Microbiol.* 64:2763-2769 (1998)) can be used to drive inducible expression. In addition, acid or alkaline-induced promoters can be used. For example, promoters that are active under the relatively acidic conditions of the vagina can be used. Alternatively, promoters can be used that are induced upon changes in the vagina in response to semen. For example, alkaline-induced promoters are used to induce expression in response to the increased alkaline conditions of the vagina resulting from the introduction of semen.

[92] A variety of signal sequences are known to direct expression of polypeptides to the membrane, extracellular space or the cell wall (*e.g.*, by covalent attachment to peptidoglycan). Exemplary signal sequences include the signal sequence from α-Amylase of *L. amylovorus* (Giraud & Cuny, *Gene.* 198:149-157 (1997)) or the signal sequence from the S-layer gene (*cbsA*) of *L. crispatus* (*e.g.*, MKKNLRIVSAAAAALLAVAPVAA or MKKNLRIVSAAAAALLAVATVSA. Signal sequences are typically located at the amino-terminus of a polypeptide.

[93] Correct localization and folding of a polypeptide can be determined using standard methods. For example, cell wall enriched fractions of *Lactobacillus* can be obtained by suspending the bacteria in a buffered, solution (*e.g.*, 25% sucrose, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) followed by treatment with cell wall degrading enzymes (*e.g.*, lysozyme and mutanolysin) and then separating out the resulting protoplasts by differential centrifugation. Fractions can then be screened by western blotting to confirm expression within the cell wall.

[94] Folding and biological activity of an expressed polypeptide can also be determined using standard methods. For example, ELISA assays using antibodies specific for the natively folded polypeptide can be used to confirm folding and three-dimensional structure of the polypeptide. Biological activity assays will of course vary depending on the activity of the polypeptide. For example, for polypeptides that bind to viral proteins, the expressed polypeptide can be tested for its ability to bind a viral protein using standard binding assays. For anti-inflammatory molecules, the expressed polypeptide can be assayed for its ability to antagonize substances that promote inflammation.

[95] When synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell. The percent deviation of the frequency of preferred codon usage for a synthetic gene from that employed by a host cell is calculated first by determining the percent deviation of the frequency of usage of a single codon from that of the host cell followed by obtaining the average deviation over all codons.

[96] The polynucleotide sequence encoding a particular polypeptide can be altered to coincide with the codon usage of a particular host. For example, the codon usage of *Lactobacillus* can be used to derive a polynucleotide that encodes a polypeptide of the invention and comprises preferred *Lactobacillus* codons. The frequency of preferred codon usage exhibited by a host cell can be calculated by averaging the frequency of preferred codon usage in a large number of genes expressed by the host cell. This analysis is preferably limited to genes that are highly expressed by the host cell. Pouwels *et al.* (*Nucleic Acids Res.* 22:929-936 (1994)), for example, provides the frequency of codon usage by highly expressed genes exhibited by various *Lactobacillus* species. Codon-usage tables are also available via the internet.

IV. Proteins of the invention

[97] The polypeptides of the invention, *e.g.*, biologically active polypeptides fused to the cell wall targeting regions of the invention) can be any polypeptide. Typically, the polypeptides of the invention are expressed under conditions to allow for biological activity of the polypeptide. In some embodiments, a disulfide bond exists in the expressed polypeptide. In some embodiments, the disulfide bond is required for the polypeptide's biological activity.

[98] Polypeptides of the invention can be of any size molecular weight. For example, the polypeptides can be between about 100 and 200,000 daltons, between about 500 and 40,000 daltons, between about 500 and 10,000 daltons, between about 10,000 and 50,000 daltons, or between about 50,000 and 200,000 daltons.

[99] Examples of classes of polypeptides that can be used according to the methods of the invention to prevent or treat pathogen infection include, *e.g.*, anti-viral polypeptides, anti-bacterial polypeptides, anti-fungal polypeptides, and polypeptides that bind to viruses, bacteria or fungi, including antibodies, antibody fragments, or single-chain antibodies.

[100] In some cases, the polypeptides of the invention will be a receptor that viral or bacterial pathogens bind to infect a host. Alternatively, the polypeptides are agents that, *e.g.*, inhibit pathogen replication, viability, entry or otherwise bind to the pathogen. In some embodiments, the polypeptides of the invention bind or inhibit sexually transmitted pathogens and other pathogens transmitted to or from the vagina. For example, since viruses require binding to a receptor on the target cell surface for infection, strategies directed at inhibiting the interaction of a virus with its host receptor are effective at preventing infection.

[101] Exemplary anti-viral polypeptides include, *e.g.*, CD4 or virus-binding fragments thereof (*e.g.*, 2D-CD4) (*e.g.*, Orloff *et al.*, *J. Virol.* 67:1461-1471 (1993)), stable CD4 trimers formed via a trimeric motif (*e.g.*, Yang *et al.*, *J. Viol.* 76:4634-4642 (2002), a dodecameric CD4-Ig fusion protein (Arthos *et al.*, *J. Biol. Chem. 277*:11456-11464 (2002)), α-defensins (*e.g.*, Zhang *et al.*, *Science* 298:995-1000 (2002), CD4 in fusion with a single chain variable region of the 17b mAb (Dey *et al.*, *J. Virol.* 77: 2859-2865 (2003))., cyanovirin-N or variants (*e.g.*, Bolmstedt *et al.*, *Mol Pharmacol.* 59:949-954 (2001); Mori *et al.*, *Protein Expr. Purif.* 26: 42-49. (2002)), herpes simplex virus entry mediator C (HveC) (*e.g.*, Cocchi *et al.*, *Proc. Natl. Acad. Sci. USA.* 95:15700-15705 (1998)), and ICAM-1. Other embodiments include, *e.g.*, viral receptors or heparin or heparin-like molecules, mannose-binding lectin, including dendritic cell-specific ICAM-3 grabbing nonintegrin (*e.g.*, Geijtenbeek *et al., Cell* 100:587-597 (2000); Feinberg *et al., Science* 294:2163-2166 (2001)), anti-HSV-1 gp120 single-chain antibody (*e.g.* Marasco *et al., Proc. Natl. Acad. Sci. USA.* 90:7889-7893 (1993); McHugh *et al.*, *J. Biol. Chem.* 277: 34383-34390 (2002)), human mAb b12, recognizing the CD4-binding site of HIV-1 gp120 (*e.g.* Saphire *et al., Science* 293:1155-1159 (2001)) or other molecules with similar specificity, including neutralizing antibodies that bind to HSV (*e.g.*, Burioni *et al., Proc. Natl. Acad. Sci. USA.* 91: 355-359 (1994)), and HIV-1 entry inhibitory protein (*e.g.*, Root *et al.*, *Science* 291: 884-888 (2001); Sia *et al.*, *Proc. Natl. Acad. Sci.* USA. 99:14664-14669 (2002)).

[102] Infection with human papillomaviruses (HPVs) is a factor that is associated with development of cervical cancer (*e.g.*, zur Hausen, *Virology* 184:9-13 (1991); Stanley, *Best Prat. Res. Clin. Obstet. Gynaecol.* 15:663-676 (2001)). Therefore, the presence of molecules that inhibit or bind to HPV is useful for preventing both HPV infection and the development of cervical cancer. Exemplary anti-HPVs polypeptides include, *e.g.* neutralizing antibodies that bind human papillomavirus type 16 E6 or E7 protein (*e.g.* Mannhart *et al., Mol. Cell Biol.* 20:6483-6495 (2000)), HPV-binding proteins, or HPV proteins that can be used to elicit an immune response directed to the virus.

[103] The capacity to bind a pathogen such as a virus or bacteria may be conferred onto the bacteria of the invention in at least several ways. The first is by making the bacteria express on its surface the normal host receptor for the virus, such as ICAM-1 for human rhinovirus HRV (major group) and CD4 for HIV. These are normal human proteins and the complete sequences of many of these genes have been determined and are stored in the database GenBank.

[104] A second method is by expressing on the bacterial surface an antibody fragment or other polypeptide that binds to a conserved determinant on the viral surface, such as VP4 on poliovirus, or gp120 on HIV. Antibody fragments (and peptides) specific for essentially any antigen can be selected, *e.g.*, from a phage-display library (Marks *et al.*, *J. Biol. Chem.* 267:16007-16010 (1992)). Antibodies can be directed to any epitope on or associated with a pathogen as well as other epitopes such as those discussed below.

[105] A third method involves the expression of carbohydrate-binding polypeptides on the surface of the bacteria. Examples of these molecules include heparin-binding polypeptides, or mannose-binding polypeptides.

[106] Anti-bacterial polypeptides include those that bind to or inhibit growth or colonization by uropathogenic *E. coli*. Exemplary anti-bacterial polypeptides include, *e.g.*, permeability-increasing protein against Gram-negative bacteria (Levy. *Expert Opin. Investig. Drugs* 11:159-167 (2002)), mammalian anti-microbial peptides, β-defensins (Ganz & Lehrer. *Pharmacol. Ther.* 66:191-205 (1995), bacteriocins (*e.g.*, Loeffler *et al.*, *Science* 294:2170-2172 (2001)) and antibodies that specifically bind to the bacteria.

[107] Anti-fungal polypeptides include those that bind to or inhibit growth or colonization by fungi such as *Candida*.

[108] Additional examples of biologically-active polypeptides useful according to the invention include therapeutic polypeptides or agents such as anti-inflammatory molecules, growth factors, molecules that bind to, or antagonize, growth factors, therapeutic enzymes, antibodies (including, *e.g.*, antibody fragments or single-chain antibodies) and molecules that inhibit or treat cancer including cervical cancer. These examples are not intended to be limiting as numerous other therapeutically active polypeptides can readily be cited.

[109] Anti-inflammatory molecules include, *e.g.*, antibodies or other molecules that specifically bind to TNF or IL-8. Other exemplary anti-inflammatory molecules include IL-10 and IL-11.

[110] Growth factors useful in the invention include, e.g., those involved in local tissue repair such as KGF, HB-EGF, FGF and TGF-β, or antagonists of these molecules.

[111] Therapeutic enzymes include, *e.g.*, nitric oxide (NO) synthase.

[112] Anti-cancer molecules include those that induce apoptosis, that regulate cell cycle such as p53, or that act as a vaccine to target cancer-specific epitopes.

[113] Vaccine molecules useful in the invention include polypeptides that elicit an immune response to viruses, bacteria, or fungi. Exemplary viral vaccines elicit response to, *e.g.*, HIV, HPV, HSV-2, or smallpox. Exemplary antigens include the glycoprotein D of HSV-2, the proteins E6 and E7 of human papilloma virus, the major outer membrane protein of *Chlamydia trachomatis* (Kim and DeMars. *Curr. Opin. Immunol.* 13: 429-436 (2001)), and aspartyl proteases of *Candida albicans* (De Bernardis *et al.*, *Infect. Immun.* 70: 2725-2729 (2002)); FimH of uropathogenic *E. coli* (Langermann *et al.*, *Science.* 276: 607-611 (1997)); IroN of extraintestinal pathogenic *E. coli* (Russo *et al.*, *Infect. Immun.* 71: 7164-9 (2003)).

V. Delivery

[114] Delivery of engineered bacteria to a desired mucosal surface depends on the accessibility of the area and the local conditions. For example, engineered bacteria may be placed in a saline solution or in a foam for delivery onto the vaginal mucosa. Foams can include, *e.g.*, one or more hydrophobically modified polysaccharides such as cellulosics and chitosans. Cellulosics include, for example, hydroxyethyl cellulose, hydroxypropyl cellulose, methyl cellulose, hydroxypropylmethyl cellulose, hydroxyethyl methyl cellulose, and the like. Chitosans include, for example, the following chitosan salts; chitosan lactate, chitosan salicylate, chitosan pyrrolidone carboxylate, chitosan itaconate, chitosan niacinate, chitosan formate, chitosan acetate, chitosan gallate, chitosan glutamate, chitosan maleate, chitosan aspartate, chitosan glycolate and quaternary amine substituted chitosan and salts thereof, and the like. Foam can also include other components such as water, ethyl alcohol, isopropyl alcohol, glycerin, glycerol, propylene glycol, and sorbitol. Spermicides are optionally included in the bacterial composition. Further examples of foams and foam delivery vehicles are described in, *e.g.*, U.S. Patent Nos. 5,595,980 and 4,922,928.

[115] Alternatively, the bacteria can be delivered as a suppository or pessary. *See, e.g.*, U.S. Patent No. 4,322,399. In some embodiments, the bacteria of the invention are

delivered in a dissolvable element made of dissolvable polymer material and/or complex carbohydrate material selected for dissolving properties, such that it remains in substantially solid form before use, and dissolves due to human body temperatures and moisture during use to release the agent material in a desired timed release and dosage. *See, e.g.*, U.S. Patent No. 5,529,782. The bacteria can also be delivered in a sponge delivery vehicle such as described in U.S. Patent No. 4,693,705.

[116] In some embodiments, the bacteria are administered orally. For example, a daily dose of about 10^8 lactobacilli can be used to restore the normal urogentital flora. *See, e.g.*, Reid *et al., FEMS Immuno. Med. Microbiol.* 32:37-41 (2001).

[117] In some embodiments, applications of engineered bacteria to a mucosal surface will need to be repeated on a regular basis; optimal dosing intervals are routine to determine, but will vary with different mucosal environments and bacterial strain. The dosing intervals can vary from once daily to once every 2-4 weeks.

[118] In embodiments where bacteriophage are introduced to transform native *Lactobacillus*, the nucleic acid of the selected bacteriophage may be manipulated such that the heterologous gene(s) replaces the genes coding for bacteriophage coat proteins, rendering the bacteriophage replication-defective. Adding these recombinant DNA molecules into cell lysates containing functional bacteriophage proteins will lead to assembly of functional bacteriophage particles carrying the heterologous gene(s). These replication-defective bacteriophage particles can then be introduced onto a desired mucosal surface to infect selected floral bacteria. The typical dosage would be 10^8 to 10^{12} PFU/ml applied to the mucosal surface. The proportion of solution to the treated surface should approximate 0.1 to 1.0 ml per square centimeter of mucosal surface. The vehicle would be similar to the vehicle described above for the bacteria.

EXAMPLE

[119] The following example is offered to illustrate, but not to limit, the claimed invention.

[120] Most viruses are transmitted through mucous membranes – nose, mouth, intestines, or genital tract. These mucous membranes are naturally colonized by vast numbers of commensal bacteria, including *L. jensenii*, *L. gasseri*, and *L. crispatus*, within the vaginal cavity of healthy women. We envision that genetically modifying *L. jensenii* to express biologically active viral binding proteins that are anchored onto bacterial surface would trap viruses within the mucosa, thus impeding the access of viruses to underlying

epithelial cells and lymphocytes. These trapped viruses may undergo an aborted infection process and/or be inactivated locally by antiviral compounds, such as lactic acid and hydrogen peroxide, secreted by the lactobacilli, thereby significantly reducing the numbers of infectious viral particles.. Accordingly, we took a modular expression approach to genetically engineered lactobacillus for surface expression of high-density HIV-binding ligand, 2-domain CD4 and cyanovirin-N. We discovered that efficient cell wall anchored display of polypeptides from 10 to 600 amino acids could be achieved by fusion to protein domains derived from native proteins of *L. jensenii*.

[121] The M6 proteins have a signature cell wall sorting signal, the LPXTG motif, followed by a stretch of hydrophobic amino acids and finally a sequence containing charged residues (KRKEEN), which serves as a critical cell surface retention signal. We initially attempted a plasmid-based modular approach to express CD4 on the surface of *L. jensenii* by utilizing two well-characterized cell-wall anchor motifs, from either the M6 protein (*emm6*) of *S. pyogenes* or the PrtP protease of *L. paracasei*, or the anchor motif from the M6 protein of *S. pyrogenes* plus an N-terminal 100-amino acid extension (CWA100) derived from the native sequence of M6 protein. Unlike the M6 protein, the sorting signal for PrtP is LPKTA. Western analysis of proteins in conditioned media and cell wall- or protoplast-associated protein pools in the modified *L. jensenii* harboring M6 or PrtP or CWA100 as cell wall anchors revealed no detectable cell wall associated 2D CD4, although abundant 2D CD4 was released into conditioned media. Flow cytometric analysis failed to identify positive surface-exposed 2D CD4.

Identification Of Putative Cell Wall Anchor Sequences

Database search of genomic sequences of *L. jensenii* allowed identification of approximately 30 contigs with putative cell wall anchor motifs. Based on a more detailed sequence homology search in the non-redundant databases available at the web site of the National Center for the Biotechnology, we selected three of these sequences, designated as C14, C191, and C370. They shared a low sequence similarity (with 23~27% identities) with Rlp of *Lactobacillus fermentum* (Turner *et al.*, *Appl. Environ. Microbiol.* 69:5855-5863 (2003)) or mucus binding protein in *L. reuteri* (Roos and Jonsson, *Microbiol.* 148:433-442 (2002)), a family of streptococcal surface proteins (Wastfelt *et al.*, *J. Bio. Chem.* 271:18892-18897 (1996)), and a cell wall-anchored proteinase in *S. thermophilus* (Fernandez-Espla *et al.*, *Appl. Environ. Microbiol.* 66:4772-4778 (2000)), respectively. All of the three sequences have LPQTG sorting signal preceding a hydrophobic region and a

charged C-terminal tail (See Figure 1). These features are common among sortase-recognized C-terminal cell wall anchor sequences in Gram-positive bacteria (Navarre and Schneewind, *Microbiol. Mol. Bio. Rev.* 63,174-229 (1999)). Among the LPXTG cell anchor motifs found in Gram-positive bacteria, only seven percent match the LPQTG sequence found in these *L. jensenii* proteins. C14, C191, and C370 proteins all contain tandem repeat domains adjacent to the cell wall anchor region, a structural feature that is frequently present in known cell wall anchored proteins (Navarre and Schneewind, *Microbiol. Mol. Bio. Rev.* 63:174-229 (1999)). The sequences of C14, C191 and C370 are displayed in Figure 2A-C.

Epitope Tagging Of Putative Cell Wall Anchor Sequences

[122] To determine the efficiency of C14, C191, and C370 to anchor heterologous fusion proteins to the cell wall of *L. jensenii*, we selected approximately 200 amino acids directly N-terminal to LPQTG sorting signal. This region, often defined as cell wall associated (CWA) domain in cell wall anchored proteins, may facilitate retention or extension of substrate sequence and thus proper proteolytic cleavage by membrane-associated sortase. To facilitate immuno-detection, c-Myc epitope (EQKLISEEDL) was fused with the N-terminus of CWA200 regions of C14, C191, and C370 in pOSEL239, 240, and 241, respectively. Western and flow cytometric analyses were employed to investigate whether the c-Myc tagged proteins were produced and targeted to the cell wall. To perform Western analyses, the modified *L. jensenii* harboring pOSEL175, 239, 240, and 241 were grown in both MRS and Rogosa SL broth to logarithmic phase. Subsequently, the cell walls were digested with mutanolysin, an *N*-acetyl muramidase that cuts the β1-4 glycosidic bond between MurNAc-GlcNAc of the glycan strands in mature peptidoglycan. Cell wall anchored proteins typically migrate as a large spectrum of fragments, following SDS-PAGE chromatography (Perry *et al.*, *J. Biol. Chem.* 277, 16241-16248 (2002)). Western analysis of proteins in cell wall enriched fractions in the bacterial cells harboring pOSEL239 (C14 anchor) and 241 (C370 anchor) revealed a ladder of c-Myc tagged proteins on reducing SDS-PAGE when the bacterial cells were cultured in both MRS and Rogosa broth (Figure 3). These patterns were absent in the cell wall enriched fraction in the bacterial cells harboring pOSEL240 (C191 anchor), demonstrating different anchoring efficiencies among LPQTG-containing sequences tested.

[123] To determine whether the Western blot positive c-Myc epitope is surface exposed in the *L. jensenii* cells harboring pOSEL239 and 241, flow cytometric analysis of the binding of anti-c-Myc antibody was performed, in reference to the bacterial

cells harboring control plasmid pOSEL175. While mean fluorescence intensity in bacterial cells harboring pOSEL239 was not distinguishable from those harboring control plasmid pOSEL175, it increased 160 fold in the bacterial cell harboring pOSEL241. While it is unclear whether steric hindrance affects the surface accessibility of c-Myc tagged CWA200 region of C14 sequence, our analysis clearly demonstrated surface exposure of the extreme N-terminus of CWA200 region of C370 sequence. This result demonstrates that this particular region of C370 can be exploited to covalently anchor heterologous peptides and proteins to the bacterial cell surface.

Surface expression of 2D CD4 on bacterial surface of *L. jensenii*

[124] We performed Western blotting and flow cytometry analyses, to determine whether 2D CD4 can be surface expressed via the CWA200 region of C14 and C370 sequences. To perform Western analysis, proteins in *L. jensenii* cells harboring pOSEL175 (control plasmid), 651 (2D CD4 plasmid without a cell anchor) (Chang *et al.*, *Proc. Natl. Acad. Sci. USA.* 100:11672-11677 (2003)), 237 (2D CD4 fused to C14 anchor), 242 (2D CD4 fused to C191 anchor), and 249 (2D CD4 fused to C370 anchor), were fractioned into cell wall enriched fractions upon cell wall digestion. In cell wall enriched protein fractions, a spectrum of higher molecular weight species were immunoreactive to pAb T4-4 in both bacterial cells harboring pOSEL237 and 249, but not in pOSEL651 (Figure 4). Such observed ladder patterns on SDS-PAGE following mutanolysin digestion resemble the patterns of known cell wall anchor proteins from bacterial surface of other Gram-positive bacteria (Perry *et al.*, *J. Biol. Chem.* 277:16241-16248 (2002)).

[125] To determine whether 2D CD4 is expressed on cell surface, the *L. jensenii* strains harboring pOSEL175, 651, 237, and 249 were probed with pAb T4-4 and subsequently analyzed for antibody binding by flow cytometric analysis. As expected, this analysis revealed indistinguishable mean fluorescence intensity in bacterial cells harboring pOSEL175 and 651. In contrast, there was significant increase in mean fluorescence intensity in bacterial cells harboring pOSEL237 and 249 relative to pOSEL175 and 651, likely as a result of covalent attachment and surface exposure of 2D CD4 molecules (Figure 5 A). To further validate the above approach, a recoded cyanovirin-N (CV-N) gene, containing *Lactobacillus*-preferred codons, was fused to the same C-terminal anchor domains that were used for successful anchoring of 2D-CD4. Flow cytometry analysis of modified *L. jensenii* harboring CV-N expression plasmids detected a 30-50 fold increase in mean fluorescence intensity relative to bacteria harboring pOSEL175 (data not shown). To

investigate the possibility that the antibody reactive CV-N molecules were surface associated via electrostatic interactions, the modified bacteria were extracted with 5 M LiCl. Flow cytometric analysis revealed indistinguishable mean fluorescence intensity in salt extracted *L. jensenii* harboring CV-N expression plasmids in reference to those washed with PBS and 2% FBS. Resistance of surface displayed CV-N molecules to extraction by 5 M LiCl reflects a behavior of covalently anchored proteins on bacterial surfaces.

[126] To address whether surface expressed 2D CD4 molecules adopt correctly folded conformation for binding gp120, additional FACS analyses were performed after bacterial cells harboring pOSEL175, 237, and 249 were probed with anti-CD4 monoclonal antibody, Sim.4, which recognizes a conformational dependent epitope. There was a significant increase in mean fluorescence intensity in the bacterial cells harboring pOSEL237 and 249 relative to pOSEL175, demonstrating that 2D CD4 were expressed in a functional form on the surface of *L. jensenii* (Figure 5B).

[127] It was unclear whether surface expression of 2D CD4 in a modular expression approach would affect expression of native cell surface associated proteins in modified *L. jensenii*. To address this issue, bacterial cells harboring pOSEL175 and 237 were probed with sulfo-NHS-biotin, and subsequently cell surface associated proteins were extracted in a buffer containing 0.4% SDS and 10 mM DTT. Western analysis of SDS-extracted proteins after probing with alkaline phosphatase conjugated avidin detected spectrum of biotinylated proteins with apparent molecular masses from 10 to > 200 kDa. The pattern of resolvable biotinylated protein species in the bacterial cells harboring pOSEL237 was similar to those in pOSEL175, indicating that native cell surface expression was not affected.

Surface expression of active 2D CD4 at wide pH range in *L. jensenii*

[128] The human vaginal cavity, when naturally colonized with lactobacillus, has a pH that varies from 3.6 to 4.5 in most women (Boskey *et al.*, *Infect. Immun.* 67: 5170-5175 (1999)), and transiently becomes neutral or weak alkaline when the male ejaculate is present. Experiments were performed to examine how pH changes would affect surface expression of an active 2D CD4 molecule in the modified *L. jensenii*. Bacterial cells were inoculated into Rogosa SL broth, either at its commonly used pH (5.4) or buffered with 100 mM HEPES, pH 7.4. The pH of the culture medium did not change substantially during active growth to OD_{600} at ~0.4. Flow cytometric analysis of binding of mAb Sim.4 to

bacterial cells harboring pOSEL237 and 249 detected significantly higher mean fluorescence intensity above control background in pOSEL175 at both pH 5.4 and 7.4. Furthermore, the level of surface-expressed CV-N remained elevated when the modified *L. jensenii* were cultured at acidic pH's that resemble those found within the human vaginal cavity (data not shown).

Lack of surface display of 2D CD4 when expressed in fusion solely via C-terminal anchor motif of 36 amino acid in length

[129] It is unclear whether a 36 amino acid C-terminal anchor motif, including LPQTG signal, a hydrophobic region, and a charged tail of C14 or C370 sequence would be sufficient to support efficient surface expression of 2D CD4 in the *L. jensenii*. To address this question, two constructs, designated as pOSEL238 harboring the C-terminal anchor motif of C14 and pOSEL262 harboring the C-terminal anchor motif of C370 were prepared and analyzed in reference to negative controls pOSEL175 and 651, and positive control, pOSEL237. Western analysis of cell wall enriched fraction in *L. jensenii* harboring pOSEL238 after probing with pAb T4-4 detected no ladder patterns resembling those in pOSEL237. Furthermore, flow cytometric analysis of mAb Sim.4 binding to bacterial cells harboring pOSEL238 failed to detect any increase in mean fluorescence intensity relative to background control in cells harboring pOSEL175 (Figure 6). Similarly, FACS analysis of the bacterial cells harboring pOSEL262, in reference to those harboring pOSEL175 and positive control pOSEL249, yielded similar negative results. Consistent with these observations, surface expression of 2D CD4 was not achieved when similar length of C-terminal anchor motifs from *S. pyogenes* and *L. paracasei* were employed. This suggests that protein sequences upstream from the characteristic LPQTG motif contribute significantly to the cell wall anchoring process and are required to display biologically active proteins on the cell wall of *L. jensenii*.

Requirement of a defined length of repetitive cell wall spanning sequence upstream of the LPQTG motif for optimal surface display of biologically proteins

[130] The native C370 sequence contains eight nearly identical tandem repeats, a characteristic of many cell wall anchor proteins in Gram-positive bacteria, in its C-terminal region upstream of the LPQTG motif (Figure 1). While two and half repeat sequences were included in the anchoring sequence of pOSEL249, it remains to be determined whether a different length of upstream sequence could be used to maximize

surface protein display. Accordingly, several constructs were prepared harboring 0, 1, 2, 4, 7, and 8 repeats of the C370 sequence. They were designated as pOSEL262, 268, 278, 280, 281, 276, respectively. To determine level of 2D CD4 molecules that adopt a correctly folded conformation, the transformed bacteria were probed with mAb Sim.4 for flow cytometry analysis (Figure 7). There was non-distinguishable mean fluorescence intensity in bacterial harboring pOSEL262 (0 repeat) from that in negative control pOSEL175, suggesting the requirement of repetitive sequence for proper surface expression of heterologous proteins. In addition, there was a significant increase in fluorescence intensity when number of repeats increased from 0 in pOSEL262 up to 3 in pOSEL278. The fluorescence intensity remained steady with additional increase in number of repeats.

Utility of native anchor sequences of *L. jensenii* in supporting surface display of proteins in a variety of lactobacillus species

[131] To determine whether the anchor sequences of C370 native to *L. jensenii* 1153 could afford protein surface display in other *L. jensenii* strains or lactobacillus species of human origin, pOSEL175 or pOSEL241, that was designed to fuse c-Myc epitope to CWA200 of C370 sequence (Figure 8A), were introduced into *L. jensenii* Xna, *L. gasseri* 1151, and *L. casei* Q by electroporation. The transformed bacteria were analyzed by Western and flow cytometric analyses, in reference to positive control *L. jensenii* 1153 harboring pOSEL241. Western analyses of cell wall digests following probing with mAb against c-Myc detected laddering patterns in transformed *L. jensenii* Xna and *L. gasseri* 1151 harboring pOSEL241 that were similar to those in *L. jensenii* 1153, and to a lesser extent in *L. casei* Q (Figure 8B). Flow cytometric analyses following immunostaining of the bacterial cells with mAb against c-Myc detected a low level of fluorescence in all lactobacillus species harboring pOSEL175 (Figure 8C), but an elevated increase in fluorescence intensity in *L. jensenii* Xna and *L. gasseri* 1151 harboring pOSEL241 as result of binding of the antibody binding to surface displayed c-Myc epitope. Additionally, there was still approximately 19 fold increase in fluorescence intensity of *L. casei* Q harboring 241 relative to that of *L. casei* harboring pOEL175. Taking these data together, the anchor sequence native to *L. jensenii* 1153 clearly exhibit a broad utility in supporting surface display of proteins in a variety of lactobacillus species, including those of human origin.

Effect of mutagenesis of LPXTG motif on surface expression of 2D CD4 in *L. jensenii*

[132] When protein A of *Staphylococcus aureus*, a well studied cell wall anchor protein, was mutated on the LPETG cell wall sorting motif, it was found that replacing amino acid proline (P) in LPQTG with amino acid asparagine (N) decreased the efficiency of protein surface display, while replace threonine (T) with serine (S) had little effect on the efficiency of protein surface display (Navarre and Schneewind, *Microbiol. Mol. Biol. Rev.* 63:174-229 (1999)). This study indicated that the P residue is probably the most important residue in LPXTG motif, and the T residue can be replaced by a similar amino acid, S. To determine whether the LPQTG motif within the C14 and C370 is indeed the critical sorting signal, the importance of P and T within the LPQTG sequence was investigated. Point mutations were generated within the LPQTG motif by PCR on both C14 and C370 sequences. The P residue was mutated to alanine (A) or asparagine (N); the amino acid T was mutated to A, S or glycine (G); the amino acid G in the LPXTG motif was mutated to A. Plasmids with the altered LPQTG motif were designated as pOSEL237P(A), pOSEL237P(N), pOSEL237T(A), pOSEL237T(G), pOSEL237T(S), pOSEL237G(A), pOSEL249P(A), pOSEL249P(N), pOSEL249T(A), pOSEL249T(G), pOSEL249T(S), and pOSEL249G(A), respectively. Western and flow cytometric analyses of the *L. jensenii* 1153 harboring the mutated constructs were performed. Compared to the *L. jensenii* harboring parental pOSEL237 and pOSEL249, those harboring pOSEL237P(A), pOSEL237P(N), pOSEL249P(A), and pOSEL249P(N) did not exhibit the characteristic higher molecular weight species spectra, upon Western blotting of cell wall enriched protein fractions with pAb T4-4. Instead, there was a marked increase in secretion of 2D CD4-CWA200 fusion protein into the conditioned medium, indicating that the 2D CD4-CWA200 fusion proteins were not covalently linked to the cell wall. A characteristic spectra of higher molecular weight species, similar to those observed with wild type pOSEL237 and pOSEL249, was evident upon cell wall digestion of *L. jensenii* harboring pOSEL237T(S) and pOSEL249T(S), suggesting that the amino acid T within LPQTG from C14 and C370 can be effectively replaced by S (data not shown).

[133] To further determine the effect of mutagenesis of LPXTG on *L. jensenii* surface protein display, the *L. jensenii* strains harboring pOSEL175, 651, 237, 249, along with the various mutant constructs, were probed with pAb T4-4 or mAb Sim.4, and subsequently analyzed for antibody binding by flow cytometry. There was a substantial decrease of mean fluorescence intensity in bacterial cells harboring pOSEL237P(A), pOSEL237P(N) compared to pOSEL237, and for pOSEL249P(A), pOSEl249P(N) comparing

to those harboring pOSEL249, indicating that there was much less 2D CD4 protein displayed on the cell surface, if any. However, the mean fluorescence intensity in the bacterial cells harboring pOSEL237T(S), pOSEL 237 (T)A, pOSEL249T(S), and pOSEL249 (T)A was comparable to *L. jensenii* harboring pOSEL237 and 249, demonstrating that replacing T with S or A has little effect on the efficiency of cell wall anchoring (Figure 9).

[134] The data from Western blot and flow cytometric analysis indicate that the amino acid P contained within LPQTG motif of C14 and C370 can not be readily substituted. In contrast, the amino acid T can be replaced with S or A, yielding a protein that still anchors efficiently to the cell wall of *Lactobacillus*.

Effect of deletion of C-terminal positive charged tail on surface expression of 2D CD4 in *L. jensenii*

[135] One of the characteristics of gram-positive cell wall anchor domains is the stretch of positive charged amino acids at the extreme C-terminus of the protein. In the M6 proteins, this sequence (KRKEEN) serves as a critical cell surface retention signal. These signature sequences have been found in other Gram-positive bacteria including *Staphlyococcus*, *Enterococcus*, *Listeria*, and *Lactobacillus* (Navarre and Schneewind, *Microbio. Mol. Biol. Rev.* 63:174-229 (1999)). Two sequences RKKRQK^{1765} and KKKRKDDEA^{1903} were identified as the positive charged tails in C14 and C370 putative anchor sequences respectively (Figure 1). To determine whether theses two sequences serve as cell surface retention signal, a series of deletion constructs were created (Figure 10). They were designated as pOSEL237-5, pOSEL237-6, pOSEL237-7, pOSEL249-8, pOSEL249-9, and pOSEL249-10, respectively.

[136] Western and flow cytometric analyses of *L. jensenii* harboring these constructs were performed. Protein species migrating at 48 kDa, representing the 2D CD4 in fusion with CWA200, can be detected by the pAb T4-4 in all the *L. jensenii* harboring the charged-tail knockout constructs, following SDS-PAGE. The secreted proteins were more abundant in *L. jensenii* cells harboring pOSEL237-5, pOSEL237-6, pOSEL237-7, pOSEL249-8, pOSEL249-9, and pOSEL249-10 than the cells harboring the parental pOSEL237 and 249. Western analysis of the proteins in the cell wall enriched fractions from all of the deletion mutants failed to detect the characteristic ladder patterns that were observed in *L. jensenii* harboring pOSEL237 or 249 (data not shown). These data suggested that the 2D CD4-CWA200 fusion proteins were not covalently linked to the cell wall.

[137] Flow cytometric analysis of modified *L. jensenii* following probing with anti-CD4 pAb T4-4 or mAb Sim.4 detected a marked decrease of mean fluorescence intensity in the bacterial cells harboring these mutant plasmids relative to those harboring parental pOSEL237 or 249 (Figure 11). These data demonstrated conclusively that deletion of the positively charged C-termini of both C14 and C370 inhibited their ability to anchor to the cell wall and display heterologous proteins.

Flexibility of LPQTG motif as a cell wall anchor signal

[138] While most cell wall anchored proteins from Gram-positive bacteria share the same sorting signal LPXTG, some of the proteins, however, have different motifs. The sorting signal for PrtP of *L. paracasei*, for example, is LPKTA (Holck and Naes. *J. Gen. Microbiol.* 138:1353-1364 (1992)). Protein L and the human serum albumin binding protein of *peptostreptococcus magnus* share a motif of LPXAG (de Château & L. Björck. *J. Biol. Chem.* 269:12147-12151 (1994); Keller *et al.*, *EMBO J.* 11:863-874 (1992); Murphy *et al.* *DNA Seq.* 4: 259-265 (1994)). When LPQTG mutated to LPQAG or LPQSG in C14 or C370 anchor proteins, there was only a slight decrease in surface display of 2D CD4, as measured by flow cytometry or Western blotting following SDS-PAGE. However, these sequences alone are insufficient to anchor proteins to the cell wall of vaginally derived lactobacilli as based on the following evidence: 1) the 36-amino acid C-terminal anchoring domain alone did not anchor c-Myc epitope, or 2D CD4 to the cell surface, 2) the prototypical M6 cell wall anchor sequence (encoded by the *emm6* gene of *S. pyogenes*) did not anchor heterologous proteins to the cell wall of vaginally derived lactobacilli, even when upstream sequences of up to 200 amino acids are included (we found a similar result when using the LPXTA motif from *L. paracasei*, and 3) the C191 protein was not an efficient anchor. These findings demonstrate that other upstream sequences contained within the CWA200 region of C14 and C370, also contribute significantly to the cell wall anchoring process.

Enhancement of 2D CD4 biological activity when fused with CWA200 of C14 and C370

[139] In order to assess biological activity, the 2D CD4-CWA200 of C14 and C370 proteins released from *L. jensenii* 1153 harboring pOSEL237-7 and pOSEL249-10 were analyzed by CD4 ELISA. The bacterial cells harboring pOSEL651, pOSEL237-7, and pOSEL249-10 were grown in Rogosa SL broth to different cell densities. Then, the cell-free conditioned media were harvested. At $OD_{600} = 0.8$, there was similar amount of 2D CD4 from pOSEL651 and 2D CD4-CWA200 from pOSEL237-7 or 249-10 released into the

medium as measured by Western blot. Nevertheless, the 2D CD4-CWA200 released from the bacterial cells harboring pOSEL237-7 and pOSEL249-10 exhibited about 2-3 fold of more activity when compared to the 2D CD4 protein from those harboring pOSEL651. The fusion of CWA200 region of C14 or C370 to 2D CD4 appeared to enhance the biological activity of the protein, probably by assisting the protein folding process. This same finding has been confirmed using a gp120 binding assay (data not shown). Western blot analysis of these proteins suggests that 2D CD4-CWA200 is significantly more stable than 2D CD4, perhaps contributing to its enhanced biological activity.

MATERIALS AND METHODS

Bacterial strains and culture

[140] Human vaginal strains of *L. jensenii*, *L. crispatus*, *L. gasseri* and *L. casei* were isolated by bacterial culture of vaginal samples obtained from healthy women. The bacterial strains were genotyped against DNA sequence of reference strains held in GenBank after amplification of 16S-23S intergenic spacer region using two primers specific to lactobacilli rRNA (Tannock *et al. Appl. Environ. Microbiol.* 65:4264-4267 (1999)). The strains were routinely grown in MRS or Rogosa SL broth (Difco, Detroit, MI) or on MRS agar plate at 37°C and 5% CO_2.

Isolation of the genomic DNA of *Lactobacillus jensenii* 1153

[141] Chromosomal DNA of *L. jensenii* 1153 was isolated based on modifications of a protocol that previously used to isolate chromosomal DNA from *L. crispatus* JCM 5810 (Sillanpaa *et al.*, *J. Bacteriol.* 182:6440-6450 (2000)). *L. jensenii* bacteria were grown in 200 ml of MRS medium at 37°C and 5% CO2 to an optical density at 600 nm of 1.0 (OD_{600} = 1.0). The cells were harvested by centrifugation at 6,600 x *g* for 10 min, and washed once in 25 mM Tris-HCl, pH 8.0, 10 mM EDTA, 50 mM glucose, and suspended after additions of 2.5 ml of 20 mM Tris, pH 8.0, 5 ml of 24% polyethylene glycol 8000, and 2.5 ml of lysozyme (4 mg/ml, Sigma Chemical Co., St. Louis, Mo) per 100 ml of bacterial culture. The resulting cell suspensions were incubated at 37°C for 1 hr. Upon addition of 5 ml of 0.2 M EDTA, the cells were centrifuged at 1,000 x *g* for 10 min at 4°C and resuspended in 10 ml of 20 mM Tris, pH 8.0 containing 50 µl of mutanolysin (15,000 U/ml; Sigma Chemical Co.). After incubation at 37°C for 1 hr, the cells were lysed by addition of 1.5 ml of 9% Sarkosyl (Sigma Chemical Co.) and 3 ml of 5 M NaCl. The cell

lysate was then mixed with 2.9 ml of 5 M sodium perchlorate. Chromosomal DNA was extracted with 17.5 ml chloroform-isoamyl alcohol (24:1 v/v) and precipitated by ethanol, air dried, and resuspended in 100 mM Tris-HCl, pH 8.0, 1 mM EDTA at a concentration of 1.5 mg/ml. Finally, the genomic DNA preparations were treated with DNase-free RNase.

Construction of *L. jensenii* genomic libraries

[142] Genomic DNA of *L. jensenii* 1153 was mechanically sheared to the desirable size range using HydroShear (GeneMachines, San Carlos, CA). Sheared DNA fragments were blunt ended by T4 DNA polymerase and Klenow enzyme, and the DNA fragments at 3 and 8 Kb were then isolated after agarose gel electrophoresis and purified using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). The resulting DNA fragments were ligated into pUC18 vector and transformed into *E. coli* DH10B cell (Invitrogen, Carlsbad, CA) to make 3- and 8-Kb genomic libraries. The bacterial transformants were selected on LB plates in the presence of X-gal and resulting colonies were arrayed into 96-well plates using a Q-pix robot (Genetix Ltd., UK). The quality of the libraries was determined by testing a plate consisting of 96 clones for uniformity of insert size and percentage of non-recombinants. Both libraries contained less than 5 % of non-recombinants and over 90% of the insert were within 20% of the expected size.

***L. jensenii* genome sequencing and assembly**

[143] The *L. jensenii* genome sequence was determined by using the whole-genome shotgun approach. Plasmid DNA of selected clones from genomic libraries was purified by either magnetic beads or the rolling circle method and sequenced from both ends using ABI BigDye terminator kits (Applied Biosystems, Foster City, CA). All sequencing reactions were run on an ABI PRISM 3700 automated DNA sequencer (Applied Biosystems). A total 15,360 sequence reads, or 160 sequence plates, were run to provide 3-fold coverage of the *L. jensenii* 1153 genome. A sequencing read is only considered successful when it generates over 50 bases of Q20 (1 possible error in 100 bases) or meets higher accuracy. The sequence chromatographs were automatically transferred to a UNIX system for base calling and quality assessment using Phred (Ewing *et al.*, *Genome Research* 8:175-185 (1998)). The pass rate is more than 80% and the average read length is in the range of 400-500 bases. The sequence assembly was performed using the Paracel GenomeAssembler or CAP4 (Paracel, Inc., Pasadena, CA). A total 484 contigs were assembled.

Identification of protein sequences with cell wall anchor motif in *L. jensenii* 1153 genome

[144] Cell wall anchored proteins of Gram-positive bacteria have a conserved C-terminal LPXTGX motif (Fischetti *et al.*, *Mol. Microbiol.* 4:1603-1605 (1990)). This hexapeptide is followed by a hydrophobic stretch of amino acids and a short charged tail, also known as a stop transfer sequence. (Schneewind *et al.*, *Cell* 70:267-281 (1992). In addition, another unique LPXTA sorting motif was identified in *Lactobacillus paracasei* (Holck and Naes., *J. Gen. Microbial.* 138:1353-1364 (1992)). To identify native cell wall anchor sequences, a computer script was written to identify motifs similar to LPXTG and LPXTA in all reading frames of the assembled contigs (resulting from estimated 75% complete genome sequence of *L. jensenii* 1153). The resulting contigs with putative cell wall anchor motifs were further verified by BLAST search for sequence homology to cell wall-anchored proteins in Gram-positive bacteria.

Construction of shuttle vector

[145] The primary shuttle vector used in these studies was pOSEL175, a modified version of pLEM7 (Fons *et al.*, *Plasmid* 37:199-203 (1997). The partial IS element was deleted by first cutting with *Sma* I, partially digesting with *Nde* I, blunting with Klenow fragment and then religating. Finally, the plasmid was subjected to site-directed mutagenesis to remove two *Mfe* I sites within the *erm* gene of pOSEL144 (Chang *et al.*, *Proc. Natl. Acad. Sci. USA.* 100: 11672-11677 (2003)). The resulting pOSEL175 plasmid has both replication origins in *E. coli* (ColE1) and *Lactobacillus* (*repA*), and thus contains the backbone of shuttle vectors used for the expression of heterologous proteins in a variety of *Lactobacillus* species.

Construction of expression cassettes in *L. jensenii*

[146] To conveniently surface anchor proteins in *L. jensenii*, an expression cassette was constructed and sub-cloned into the *Sac*I and *Xba*I sites of pOSEL175. The cassette contains four components, including a lactobacillus-compatible P_{23} promoter, CbsA signal sequence of *L. crispatus*, DNA encoding a heterologous protein, and covalent cell wall anchoring domains from known or putative cell surface proteins in Gram-positive bacteria. Our detailed analyses of constructs harboring a series of promoters and signal sequences indicated that a combination of the P_{23} promoter from *Lactococcus lactis* (van der Vossenet *et al.*, *Appl. Environ. Microbiol.* 53:2452-2457 (1987)) and the signal sequence from the CbsA

of *L. crispatus* (CbsAss) drives the highest levels of protein expression of 2D CD4 in the construct designated as pOSEL651 (Chang *et al.*, *Proc. Natl. Acad. Sci. USA*. 100:11672-11677 (2003)). Unique restriction sites, including *Sac*I, *Eco*RI, *Nhe*I, *Mfe*I, and *Xba*I were placed between each component from 5' to 3' ends, respectively. Amplification of each component by PCR was performed using *Pfu* DNA polymerase. Oligonucleotide primers for PCR amplification of various portions of the fusion constructs detailed in this study include the following:

P23.f	5'-GTGGAGCTCCCCGAAAAGCCCTGACAACCC-3'
P23.r	5'-GGAAACACGCTAGCACTAACTTCATT-3'
2DCD4.f	5'-GCGGCTAGCAAGAAAGTTGTTTTAGGTAAA-3'
2DCD4.r	5'-GCACAATTGTGATGCCTTTTGAAAAGCTAA-3'
CbsAss.f	5'-GCGAATTCAAGGAGGAAAAGACCACAT-3'
CbsAss.r	5'-1CCAGCTAGCTGAAACAGTAGAAACGGC-3'

[147] Proteins designed for surface expression include a 10-amino acid c-Myc peptide (EQKLISEEDL) and the first 183 residues comprising the N-terminal two extracellular domains of human CD4 (2D CD4). The 2D CD4 protein was recoded to conform to a preferred lactobacillus codon usage. All expression constructs were confirmed by DNA sequence analysis prior to transformation into *L. jensenii*.

Construction of c-Myc fusion to putative cell wall anchor sequences of *L. jensenii*

[148] We chose initially epitope tagging to determine the level of protein expression and whether it is feasible to use a defined length of putative cell wall anchor sequence for surface display of biologically active proteins. In order to not disrupt functioning of C-terminal sorting motif, oligonucleotide primers containing the 10 amino acid c-Myc epitope (EQKLISEEDL) in the 5' end were designed, allowing fusion of c-Myc epitope to the N-terminus of the putative cell wall anchor sequences, including C14, C191, and C370 from the genome of *L. jensenii* 1153. The c-Myc sequences were either fused directly to the cell wall anchor motif of these proteins (the C-terminal 30 amino acids of C14, C191, and C370) or to sequences containing the C-terminal cell wall anchor domain and various lengths of contiguous upstream amino acids. Most notably, c-Myc was fused to a 200 amino acid sequence containing the cell wall anchor domain and upstream amino acids (designated CWA 200).

Myc14nhe (5' primer)

(GCGCTAGCGAACAGAAACTGATCTCCGAAGAGGACCTGGTAACTC GTACTATCAATGTA)

Myc14mfe (3' primer)

(CGCCAATTGCTACTTTTGACGTTTCTTTCT)

Myc191nhe (5' primer)

(GCGCTAGCGAACAGAAACTGATCTCCGAAGAGGACCTGGACGTAG TAATTCCAGGAA)

Myc191mfe (3' primer)

(GCGCAATTGTTAATCTTCTTTTCTCTTCTT)

Myc370nhe (5' primer)

(GCGCTAGCGAACAGAAA CTGATCTCCGAAGAGGACCTGTTGAAGAAGGCAGAAGAAGT)

Myc370mfe (3' primer)

(CCGCAATTGTTATGCTTCATCATCTTTTCT)

[149] All of the PCR products with expected size were gel-purified and digested with both *Mfe*I and *Nhe*I. The resulting fragments were ligated with *Mfe*I/*Nhe*I double digested pOSEL651 to make c-Myc fusion in pOSEL239 (via CWA200 of C14 sequence), pOSEL240 (via CWA200 of C191 sequence), and pOSEL241 (via C370 sequence), respectively. The resulting plasmids were electroporated into *L. jensenii* 1153.

Subcloning of cell wall-anchoring sequences into shuttle vector

[150] Three putative surface proteins containing C-terminal LPQTG anchoring motif were chosen to determine their ability to express foreign proteins on the cell wall of *L. jensenii* 1153. The DNA regions containing the C-terminal LPQTG domain and their upstream 200 amino acids of these surface proteins (tentatively designated as CWA200 region) were amplified by three sets of primers, as described below,

C14: 5' primer (GCGCAATTGGTAACTCGTACTATCAATGTA)

3' primer (CGCTCTAGATACACAAACTATTTTACGGTC)

C191: 5' primer (GCGCAATTGGACGTAGTAATTCCAGGAACA)

3' primer (CGGTCTAGACCAAGCAATTTATATATTGCT)

C370: 5' primer (GCGCAATTGAAGAAGGCAGAAGAAGT)

3' primer (CCGTCTAGATTATGCTTCATCATCTTTTCT)

[151] The internal *Mfe*I site of C14 anchor domain and the internal *Xba*I site of the C370 domain were mutated by site-directed mutagenesis before enzymatic restriction. All the PCR products of predicted size were gel-purified and digested with both *Mfe*I and *Xba*I. The resulting fragments were ligated with *Mfe*I/*Xba*I double digested pOSEL651, which contains P23-regulated secreted 2D CD4, to make plasmid pOSEL237 (via CWA200 of C14 sequence), pOSEL242 (via CWA200 of C191 sequence) and pOSEL249 (via CWA300 of C370 sequence), respectively. Alternatively, the C-terminal 36-amino acid anchor motif of C14 sequence was similarly cloned into shuttle vector by using following two primers.

Mfec14up: 5' GCGCAATTGCCACAAACTGGTTCTAAGACT

Xnac14lo: 3' primer (CGCTCTAGATACACAAACTATTTTACGGTC)

[152] All of the resulting plasmids after verification of DNA sequences were electroporated into *L. jesneii*, *L. gasseri,* and *L. casei*.

Subcloning of the repetitive cell wall spanning regions of C370 sequence

[153] Different repetitive cell wall spanning regions upstream the C370 LPQTG motif were amplified from the genomic DNA of *L. jensenii* 1153. The same 3' primer (5'-CCGTCTAGATTATGCTTCATCATCTTTTCT-3') was used, in pair with the following 5' primers for each PCR reaction.

Zero repeat: 5'-CGGCAATTGCCTCAAACTGGTACTGA-3'

One repeat: 5'-CGGCAATTGGGTCAAACTACAAATAAAGAT-3'

Two repeats: 5'-CGCCAATTGGGTCAAACTACTGATAAGAGT-3'

Three repeats: 5'-GCGCAATTGGGTCAAACTACAAATAAAGAT-3'

Four-eight repeats: 5'-CGGCAATTGGGTCAAACTACTGACAAGAGC-3'

Both *Mfe*I and *Xba*I sites in these primers are underlined.

[154] All the PCR products of predicted size were gel-purified and digested with both *Mfe*I and *Xba*I. The resulting fragments were ligated with *Mfe*I/*Xba*I double digested pOSEL237, which contains P23-regulated secreted 2D CD4, to make plasmid pOSEL262 (with no repeat), pOSEL268 (with one repeat), pOSEL278 (with two repeats), pOSEL284 (with three repeats) pOSEL280 (with four repeats), pOSEL275 (with six repeats), pOSEL281 (with seven repeats) and pOSEL276 (with eight repeats), respectively.

Bacterial Transformation

[155] Plasmids were introduced by electroporation into *E. coli* DH12S (Invitrogen). For shuttle plasmid construction and maintenance, the transformed *E. coli* DH12S cells were grown in LB broth (Difco) at 37°C, supplemented with 100 μg/ml ampicillin or 300 μg/ml erythromycin. After DNA sequence verification, *E. coli*-derived plasmids were transformation into *L. jensenii*, *L. gasseri*, and *L. casei* according to Luchansky et al (*J. Dairy Sci.* 74, 3293-3302 (1991)) with modifications. Briefly, freshly inoculated *L. jensenii* were cultured in MRS broth to 0.6-0.7 at OD_{600} at 37°C and 5%CO_2. The bacterial cells were harvested, washed and re-suspended in 952 mM sucrose and 3.5 mM $MgCl_2$ at 4°C. Using a pre-chilled 0.2 cm gap cuvette, competent cells were added with 1~2 μg of DNA and electroporated immediately at 2.5 kV/cm and 200 ohms using Gene Pulser II (Bio-Rad, Hercules, CA). Afterward, cells were allowed to recover in pre-warmed MRS broth for two hours at 37°C, prior to being plated on selective MRS agar plates containing 20 μg/ml erythromycin, a concentration also used for routine propagation of transformed *L. jensenii* in liquid media.

Site-directed mutagenesis of LPXTG motif of putative cell wall anchor sequences

[156] Point mutations were generated using QuickChange® XL Site-Directed Mutagenesis Kit from Stratagene (La Jolla, CA). Plasmid pOSEL237 (expressing 2D CD4 anchored via CWA200 of C14 sequence) and plasmid pOSEL249 (expressing 2D CD4 anchored via CWA200 of C370 sequence) were used as templates. The mutagenic primers were designed based on the nucleotide sequences corresponding to LPQTG and its flanking sequences on C14 and C370:

[157] C14-GAAAGTAAGAAGACTTTACCACAAACTGGTTCTAAGACTGAA

[158] C370-CATAAGCAAACTCTATTGCCTCAAACTGGTACTGAAACTAACCCAC

[159] The replacement nucleotides were selected using *L. jensenii* 1153 preferred codons:

237P(A): Proline on LPQTG of C14 was replaced with Alaine

5'-GAAAGTAAGAAGACTTTAGCACAAACTGGTTCTAAGA-3'

5'-GTCTTAGAaccAGTTTGTGCTAAAGTCTTCTTACTTTC-3'

237P(N): Proline on LPQTG of C14 was replaced with asparagine

5'-GAAAGTAAGAAGACTTTAAATCAAACTGGTTCTAAGAC-3'

5'-GTCTTAGAACCAGTTTGATTTAAAGTCTTCTTACTTTC-3'

237T(A): Threonine on LPQTG of C14 was replaced with Alanine

5'-AGAAGACTTTACCACAAGCTGGTTCTAAGACTGAAC-3'

5'-GTTCAGTCTTAGAACCAGCTTGTGGTAAAGTCTTCT-3'

237T(G): Threonine on LPQTG of C14 was replaced with Glycine

5'-AGAAGACTTTACCACAAGGTGGTTCTAAGACTGAAC-3'

5'-GTTCAGTCTTAGAACCACCTTGTGGTAAAGTCTTCT-3'

237T(S): Threonine on LPQTG of C14 was replaced with Serine

5'-AGAAGACTTTACCACAAAGTGGTTCTAAGACTGAAC-3'

5'-GTTAGTTTCAGTACCACTTTGAGGCAATAGAGTTTG-3'

237G(A): Glycine on LPQTG of C14 was replaced with Alanie

5'-GACTTTACCACAAACTGCTTCTAAGACTGAACAAG-3'

5'-CTTGTTCAGTCTTAGAAGCAGTTTGTGGTAAAGTC-3'

249P(A): Proline on LPQTG of C370 was replaced with Alaine

5'-CATAAGCAAACTCTATTGGCTCAAACTGGTACTGAAAC3'

5'-GTTTCAGTACCAGTTTGAGCCAATAGAGTTTGCTTATG-3'

249P(N) Proline on LPQTG of C370 was replaced with Asparagine

5'-CATAAGCAAACTCTATTGAATCAAACTGGTACTGAAAC3'

5'-GTTTCAGTACCAGTTTGATTCAATAGAGTTTGCTTATG-3'

249T(A) Threonine on LPQTG of C370 was replaced with Alanine

5'-CAAACTCTATTGCCTCAAAGTGGTACTGAAACTAA-3'

5'-GTTAGTTTCAGTACCAGTTTGAGGCAATAGAGTTTG-3'

249T(G) Threonine on LPQTG of C370 was replaced with Glycine

5'-CAAACTCTATTGCCTCAAGGTGGTACTGAAACTAAC-3'

5'-GTTAGTTTCAGTACCACCTTGAGGCAATAGAGTTTG-3'

249T(S) Threonine on LPQTG of C370 was replaced with Serine

5'-CAAACTCTATTGCCTCAAAGTGGTACTGAAACT-3'

5'-GTTAGTTTCAGTACCACTTTGAGGCAATAGAGTTTG-3'

249G(A) Glycine on LPQTG of C370 was replaced with Alanine

5'-CTCTATTGCCTCAAACTGCTACTGAAACTAACCCAC-3'

5'-GTGGGTTAGTTTCAGTAGCAGTTTGAGGCAATAGAG-3'

[160] Polymerase chain reaction (PCR) cycling conditions were 95 °C for 50 sec, 60°C for 50 sec, and 68°C for 12 min for a total of 16 cycles.

[161] *Dpn* I enzyme were added to the amplification mixture after the PCR reaction to degrade the parental plasmids. Newly synthesized plasmids were introduced into

chemically competent *E. coli* Top10 cells (Invitrogen) following the manufacturer's recommendations. Plasmids were maintained and amplified in LB broth (Difco) supplemented with 200 μg/ml erythromycin. After DNA sequence verification, *E. coli*-derived plasmids were transformation into *L. jensenii* according to Luchansky *et al* (*J. Dairy Sci.* 74, 3293-3302 (1991)) with modifications. MRS containing 20 μg/ml erythromycin was used for selection and propagation of transformed *L. jensenii* containing the mutagenic plasmids.

Deletion analysis of positive charged C-terminal sequences of putative cell wall anchor proteins

[162] A series of deletion mutants, in which positively charged amino acid located at the C-terminus of C14 and C370 were generated by PCR amplification. Plasmids pOSEL237 and pOSEL249 were used as template. An oligonucleotide complementary to 2D CD4 sequence on pOSEL237 and pOSEL249 (CD4F 5'-GATCGTGCTGATTCACGTCGT-3') was used as forward primer. The following oligonuclotides (with restriction sites underlined) were used as reverse primers for amplifying the C-terminal of 2D CD4 cDNA and complete C14 and C370 CWA200 sequences:

C14-7	5'-GCGCTCTAGACTAAACACCTAAGCCTAATAAGC-3'
C14-6	5'-GCGCTCTAGACTAGTTAACACCTAAGCCTAATAAG-3'
C14-5	5'-GCGCTCTAGACTATCTGTTAACACCTAAGCC-3'
370-10	5'-GCGCTCTAGATTAAAAAATTCCTGCGCCTAATG-3'
370-9	5'-GCGCTCTAGATTATGCAAAAATTCCTGCGCCTAATG-3'
370-8	5'-GCGCTCTAGATTACTTTGCAAAAATTCCTGCGCC -3'

[163] All reverse primers contained a *Xba*I restriction site. The cycling conditions were 94 °C for 45 sec, 60°C for 45 sec, and 72 °C for 90 sec for a total of 18 cycles. The PCR products were gel-purified and digested with both *Mfe*I and *Xba*I, and then sub-cloned into *Mfe*I/*Xba*I double digested pOSEL237 and pOSEL249, respectively. The sequences were verified by nucleotide sequencing, and the constructs were electroporated into *L. jensenii* for protein analysis.

Western Analysis of heterologous protein expression in *L. jensenii*

[164] Genetically modified *L. jensenii* cells were grown in Rogosa SL broth buffered with 100 mM HEPES, pH 7.1 at 37°C and 5% CO2. To determine level of soluble

proteins, conditioned media were collected after centrifugation at 12,000 x g and proteins were then precipitated with TCA at a final concentration of 20%. TCA precipitates were washed with ethanol, air dried and heat denatured in 50 mM Tris-HCl, pH 6.8, 0.4% SDS, 6% sucrose, 10 mM dithiothreitol, and 0.01% bromphenol blue (1x reducing SDS-PAGE buffer). To determine relative amounts of cell-associated proteins in *L. jensenii*, bacterial cells were extracted without inducing cell lysis in 100 µL per OD600 unit of 1 x SDS-PAGE buffer at 37°C for 30 min. Extracted proteins were harvested following centrifugation at 12,000 x g for 5 min and subsequently heat denatured. Soluble proteins were separated from bacterial cells by centrifugation at 14,000 x g and resolved by SDS-PAGE in a 4-12% NuPAGE system (Invitrogen) in the presence of antioxidant according to manufacture's recommendation. After electrophoretic separation, proteins were electroblotted on to polyvinylidine difluoride membranes (Millipore) in 20% methanol, 20 mM Tris, and 50 mM glycine. The blot was then probed with polyclonal rabbit anti-CD4 antibodies, T4-4 (the NIH AIDS Research and Reference Reagent Program) or rabbit anti-CV-N pAb, and monoclonal antibody against c-Myc (Invitrogen). The antigen-antibody reaction was visualized by using chromogenic detection reagents (Promega, Madison, WI) for alkaline phosphase conjugated anti-rabbit IgG (for CD4 detection) or enhanced chemilluminescent reagents (Amercham Biosciences, Piscataway, NJ) for horseradish peroxidase (HRP) conjugated anti-mouse IgG (for c-Myc detection). Similarly, level of c-Myc tagged proteins were probed with mAb against c-Myc (Invitrogen) and bound antibodies were detected with HRP-conjugated anti-mouse secondary antibodies (Amersham Biosciences).

Enzymatic digestion of *L. jensenii* cell wall by muramidase

[165] Bacterial cultures containing 10^9 bacteria were centrifuged at 12,000 x *g* for 5 min. The resulting cell pellets were washed once in 20 mM HEPES, pH 7.2 and suspended in 100 µL of 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 25% sucrose (Piard *et al.*, *J. Bacteriol.* 179:3068-3072 (1997)). The bacterial cell walls were digested in the presence of muramidase, mutanolysin (Sigma Chemical Co.) at a final concentration of 15 units/ml for 1 hr at 37°C. Afterward, the cells were centrifuged at 2,500 x *g* for 10 min to isolate cell wall enriched fraction from protoplast-enriched one. The resulting samples were heat denatured after addition of 25 µl of 4 x or 125 µl of 1x reducing SDS-PAGE buffer to cell wall or protoplast enriched fractions, respectively. Alternatively, CD4 ELISA was used to analyze proteins in the cell wall enriched fractions without additional sample treatment.

Labeling of surface exposed proteins in *L. jensenii* with sulfo-NHS-biotin

[166] Surface exposed lysyl residues of surface proteins in *L. jensenii* were probed by use of membrane impermeable sulfo-*N*-hydroxysuccinimido (NHS)-biotin. Surface labeling of Gram-negative bacteria *Helicobacter pylori* by NHS-biotin allows identification of genuine cell surface proteins (Sabarth *et al.*, *J. Biol. Chem.*70:27896-27902 (2002)). Approximately 10^9 of *L. jensenii* bacteria at log phase were washed once and suspended in PBS. Sulfo-NHS-biotin was added to 1 ml of cell suspension at a final concentration of 1 mM and allowed to incubate for 30 min at room temperature, with a continuous rotation. Afterward, the biotinylation reaction was quenched with addition of 50 mM Tris, pH 8.0, and the cells were washed once with 20 mM HEPES, pH 7.2. The cell-associated proteins were extracted without inducing cell lysis in 125 µl of 0.4% SDS, 6% sucrose, 10 mM DTT, 50 mM Tris-HCl, pH 6.8 at 37°C for 30 min. The extracted proteins were separated from bacterial cells by centrifugation at 14,000 x *g* for 5 min. After heat denaturation, proteins were resolved in a 4-12% NuPAGE (Invitrogen). Biotinylated proteins and their mobility shift were determined, following probing with alkaline phosphatase conjugated strepavidin or other immunological probes.

Analysis of surface expression of 2D CD4 by flow cytometry

[167] Transformed *L. jensenii* harboring plasmids for surface protein expression or protein secretion in pOSEL651 were in grown in MRS broth in the presence of 20 µg/ml erythromycin at 37°C and 5% CO2 for overnight (with OD600 > 3). The overnight cultures were then sub-cultured at 1:50~100 dilutions in erythromycin-containing MRS or Rogosa SL Broth that is buffered with 100 mM HEPES, pH 7.1 except otherwise indicated. One ml of cell cultures at OD_{600} ~= 0.4 was centrifuged at 12,000 x g for 5 min. The resulting cell pellets were washed twice and suspended in 1x PBS containing 2% FBS. Afterward, cells were surface-stained in 2% FBS in 1x PBS for 30 min by using specific antibodies (1:1000 dilution for rabbit polyclonal T4.4 or 50 µg/ml for monoclonal Sim.4 per 2×10^8 cells,), followed by FITC or phycoerythrin-conjugated anti-rabbit or mouse antibodies (Becton-Dickinson, Mountain View, CA). A similar protocol was developed for the detection of surface expressed CV-N. Controls consisted of isotype-matched monoclonal antibodies (Becton Dickinson). Labeled cells were fixed with 1% (v/v) paraformaldehyde and analyzed in a FACScalibur system (Becton-Dickinson) running with the CellQuest

software. Density plot output (Side scatter or forward scatter vs fluorescence) in background control was obtained from *L. jensenii* harboring pOSEL 175. The shift in mean fluorescence intensity between the plots was taken as a measure of antibody binding to bacterial surface and calculated using FLOWJO software.

Enzyme-linked immunosorbent assay

[168] The concentration of correctly folded 2D CD4 proteins was determined by CD4 capture enzyme-linked immunosorbent assay (ELISA) that was modified according to McCallus *et al.* (*Viral Immunol.* 5:209-219 (1992)). 2D CD4 proteins with correct conformation in bacteria-free conditioned media were captured on a MaxiSorp 96-well plate (Nalge Nunc International, Denmark) by monoclonal antibody Sim.4 at 2.5 μg/ml. After washes in 1x Tris-buffered saline containing 0.05% Tween 20, the bound CD4 molecules, in reference to *E. coli* derived and refolded 2D CD4 standards, were probed with rabbit polyclonal antibodies, T4-4, then detected by horseradish peroxidase-conjugated anti-rabbit IgG (Amersham Biosciences) in the presence of 3,3',5,5' tetramethylbenzidine (Neogen Corp., Lexington, KY) at room temperature and in the dark for 30 minutes. The reaction was stopped after addition of 100 μl of 0.5 M H_2SO_4 and absorbance at 450 nm was read using microplate reader (Molecular Devices, Sunnyvale, CA).

[169] The above example is provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, databases, Genbank sequences, patents, and patent applications cited herein are hereby incorporated by reference.

SEQUENCE LISTING

SEQ ID NO:1

C14 SEQUENCE

[1]MNDSSIGTINITNDITITGKVNGLTTSGISDINKHFLYLQSEGSARDLTINGNGHRINFA
GYSIALQNKNYTNAANPWNITLKDMTIEGSKYDYSPISFYGRKSNTENSKLTFDGVT
ANLNDRPLVDKYGENLPVHFAGENNITLNNMSIGYNLVTGKTVKFDSGNTTFNVDG
KVTGNSINPDNWVIRSTENASNSENPSTLINEGATVTINAKSDDLRGIYAGRQLTAGQ
PIYGVTVINGTLNAKMAAGHSTAIWSHDLEIGKKGNVTIHTKQTNQADGVENGTSNS
VTNYNGTHYAPISLGVGPISSVASPLSKQTVSLINNGSLTIIRDTAKKTLVPLISMGDGS
LSSNTTLKFSVGAGATLDLQDKAGTFRYGIEPSTPLNGLVTLWGTSGTDLLEFLTPAY
VNLQRTGDIRGTLIRMEGVYNSTTVNGPTPVAQWDQGNKTTTPNDVWYVRYLISAN
QWGNNSGQFMGKDQHPNTVVAKKGVDTLYNSNATVLMSKNQGADKYENGTMPT
EVQQALHLNSFLNNFNFWRPQRMAMGSKLNDNPDVKIDDFDKYHAEAQTIDGTTR
QTLSDLDANKGLKDLIGPDEQPITDFKDIVKHVTWYNSATDKDEWNKIMIQPTDSKD
PSARVPYPEPQNPTGNLKTTDGFAWAKVTYADGSVDFVKIPLKVTEKKYSEELTPSY
PGVSVEQGKSDSVDPSFKDENDKAADAPAGTKYTAGENTPDWIKVDPDTGKVTVSP
TDDTSVGSHDISVTVTYPDSSTDQLTVPVTVTEKSNLAEKYPVSYDKLNVEKPSGDT
PATGAVDPKAAADMPEGAITGYEKGDFDAPAGVTIDVNHDTGKVTASVGKNATLG
SFEVPVKVTYSDGTYAEVKVPVSITGNKVDPGSGDVVYYGDQSMVVFNGNLTTVH
KTTDSHELSAKDSAFQTITYYSDWNKKGNIVSDYNKHVIYKLSADGTKYVNEADAT
DSFDASAISFNWQKGYEVNTGVDNFSNGSADTLYQLEKGAVNSEEQTDANDPSGLA
GNSKYRYDFSISDTNVLQKLGLSPAGYNAWANVYYNFLGATGKINIPVNYGSEVSTD
EAGIKNYLATNSISGKTFVNGNPTGIKWAENGMPGKDGKFAASNMTGIVEFTFDNGT
KLNVQVTFKTGSHVSTSGSKVNDDTNLYVERTIEYDVTGTGHSPINSVTQKVHYVRD
GYHKINADGTDAGEIIWNEWKLADGQTAEFPEYSVDQITGYDAYINGAKATQVDAA
KVAETNGTPQNGQNITVTYKKQNSTPVPYKPGKDGVNDAINRYVTRTIIVKEPGKEP
QTITQTVHFTNEDKDGNSGYKDPVTGEIKYNTDWHVASDLNAKTGSWEEYTAPSVT
GYTPSQAKVEAKTVTAETEAASVTISYTKNADIPVPYKPGKDGVNDAINRYVTRTIIV
KEPGKEPQTITQTVHFTNEDKDGNSGYKDPVTGEIKYNTDWHVASDLNAKTGSWEE
YTAPSVTGYTPSQAKVEAKTVTAETEAASVTISYTKNADIPVPFDPSNKDMYREVTR
TINVVDPITGKISTSVQTAKFTREDKNSNAGYTDPVTGKTTMNPWTPAKQGLRAVNV
EQIKGYVAKVDGNVDAVVVTPDSANMVVTITYQANKPEGQNITVKKDTVPDPADGI
KNKDDLPDGTKYTWKEVPDVNSVGEKTGIVTVTFPDGTSVDVKVTVYVDPVVESN
RDTLSKEANTGNTNVAKAATVTSSKVESKKT**LPQTG**SKTEQVGILGLAIATVGSLLG
LGVNRKKRQK[1765]

SEQ ID NO:2

C191 SEQUENCE

[1]MPVANKPEGTVHTTYSWKDNIIPDTTKPGTKYGIVEVNFPDGSTKDVPVEVKVTSL
ASDYQNKIDTKQIIAKYKGNIPQASDGIANKDQATKEGDKDFPSLADVLAPNGIQWK
KNFEPDLSKPGLTSGEAILTFKDGSTAEVTIPVLVQTDADRNTPETQTIKTLPGQTVNP
EDGVINLHKPGENNPQLPDGTKVTFDNQSDVDDFTKHGMPGSDKSFDATVTYPDGT
TDKIKLPVHITADNEVNTPITQGIITPKDSVPDANKGIANLKKATTKEGKTYPALPENT
TVEWVNPGQMKTELENAKGGTTKNYDAVVIYPDKSTEIVSIPVTVATDADTYKVVT
QPIDLKDRNLPDNADDGITNLHKPADFKTPQLPDGTHAEWQDKDAAQEVVKNLKPG
ETVKLPATVVFPDGSKKGEGIDVSVHLHGQSDDYNIETQPVNTDKDGNLPENADSGI
KNLGKLPEGTHASWGDGAQDIAKNLKPGETKDVPATVVFPDGSKKEITIPVHREGQS
DGYDVEPQLVNTDKNGQLPNAKEGIKNLADLPEGTNPTWADRAQDKINKTKPGTDT

TAQVVVTFPDGSTKEVTVPVHKHGQSDDYGDKIVTQRVETDSHGQLPENADSGIKN
LGDLPEGTHAVWGQGAQTIVDGMKPGETKDVPATIEFPDGSTKDVTIPVYKTSTRDQ
GTLNPPTDKVSVDDTKHITDEDKGKVIDNVKKSNPDKDITDAHVDDDGTFHGKVDG
QDVVIPGTETVVEKQKESLNPPTDKVPVDDTKHITDEDKGKVIDNVKKSNPDKDITD
AHVDDDGTFHGKVDGQDVVIPGTETVVEKQKESLNPPTDKVPVDDTKHITDEDKGK
VIDNVKKSNPDKDITDAHVDDDGTFHGKVDGQDVVIPGIETVVEKSTNNQKSDTNK
GLISNDNSEKNSHMINANVNTKSRNSLSAKQNR**LPQTG**SETSGLSALGLAMLSLVGL
GFLIKKRKED[974]

SEQ ID NO:3

C370 SEQUENCE

[1]MFYQIDPALAPYIDKIVFSRALLSDGEATKDTSNEVPGATNVWTSGVLTTQNGPIRA
ALAGSTSSTYKIYLKADTPNSILSKPLSFTMWARYSSGHDMVSDFSKNLILNDNETTT
FSSNNFFKSLDIVNNDGPILDNMSVDYSNKTVNTRYRVNGSLLGDKSNLTLRIRGND
NLLKLIDKVKISNKTYTLANNTLKYRTGELYINDIGGSLGFLSSLSNRQDFNVTFYLK
NGKSFADALTSESQKFDFQFGIYDTTDYATAFHSLDTVTNSLSTKTYTTGDKYNNQT
YDLSTFKTILDKLIKQKQDNPTTYLSFEDKKISATENNPYEAVKLALESPTFTNISIAKS
LVNAADCKQLDNTAKWAWDNGARDDLLKYLDVATKVASYIHLEFPTKPTDFSGLL
LRYTRAGTFISAVDSDRDGVLDITEIDNSYGMNPSVYDTDGDGISDGQELREGRDPG
VAPFNWTDANGNQLSIDVDTTTISGQLGNHNYHNEVMQPRTVNLYKVDDTGKKTLI
AYTTSAVDQNGSFTLSKFTLNKGDKLVIGYVTPRTNKSLTDKDTILQQAFPTEQFSNE
IIVKGKQVTVTFNMNGVSDDENQDIKVEKDSSFNKDSLTLPTPTMKTGYSFKEWNTQ
ADGKGTVVTADTIFDTDTTVYAIGEKIKLPNPTNIKAETRTDDKTKSQETIITGKATPG
ATVTIKDNLGNEIGTGVANDAGNFEIKTTSPLAEATKVSVEATKGGESSDAVEATVE
QNNFQKGNPLIQPASPTAVTAVTIKASDGTNNSTTVTGKAAAGETVTVKDSSGNEIG
TGVVGEDGTFTITTNKPIAENERIQVVVTKDDAESEPTEAVVTAKTEPTNPTEVTAKT
LPDGNSDSTIVAGKGKAGE
VVTVKNDAGKVIGTGKVSDDGTFSIKTDEVIEPGKQVSVITTNDGMDSIPVPVTVSGE
TITSIKQSAKAAVDNLTYLNNAQKQSAKDAIDSANTVDEITTAKNNAVSTDTNMKDL
SEDTKLAADKTQDPYLNADLDKKQAYDKAVEEAQKLLNKETGTSVGADKDPAEVA
RIKQAVDDAYDALNGNSSLDDAKQKAKDAVDKNYTNLNDKQKETAKKRIDSAKSE
DEVNNADKINSGLNEKMGELKEVSNLSDKIETTSNYSNADSDKKQAYKETADKIHET
VAPSGDDLTTDDVNNLITDEATKRAALNGDAREKARQE
LENNYNSGKSLQDGSTLDPRYYNASEEKKQAFQKALDNAKKALDNSETTEAEYKSA
NDELQKAKADLDGQTTDKSKLDDAIKDANNAKGTDKYKNASDDTKSKFDEALKKA
EEVKNNSNATQKEVDDATNNLKQAQNNLNGQTTDKSKLDDAIKDANNAKGTDKY
KNASDDTKSKFDDALKKAEEVKNNSNATQKEVDDATNNLKQAQNDLDGQTTDKS
KLDEAITDANNTKLTDKYNNASDDTKSKFDEALKKAENVKNDSNATQKEVDDATN
NLKQAQNDLDGQTTDKSKLDEAITDANNTKSTDKYNNASDDTKSKFDEALKKAEE
VKNNSNATQKEVDDATNNLKQAQNNLDGQTTDKSKLDEAITDANNTKSTDKYKNA
SDDTKSKFDDALKKAEEVKNNSNATQKEVDDATNNLKQAQNDLDGQTTNKDTLND
AIKDANDAKGTDKYKNASDDTKSKLDETLKKAEEVKNNSNATQKEVDDATNNLKQ
AQNDLDGQTTDKSKLDEAIKSADDTKSTDKYNNASDDTKSKFDEALKKAEEVKNNS
NATQKEVDDATKNLKQAQNDLDGQTTNKDAINDAIKDANNAKGTDKYNNASDDT
KSKFDDALKKAEDVKNDSNANQKEVDDATKNLKNTLNNLKGQPAKKANLIASKDN
AKIHKQTL**LPQTG**TETNPLTAIGIGLMALGAGIFAKKKRKDDEA[1903]

SEQ ID NO:4

KKAEEVKNNSNATQKEVDDATNNLKQAQNDLDGQTTDKSKLDEAIKSADDTKSTD
KYNNASDDTKSKFDEALKKAEEVKNNSNATQKEVDDATKNLKQAQNDLDGQTTN
KDAINDAIKDANNAKGTDKYNNASDDTKSKFDDALKKAEDVKNDSNANQKEVDD
ATKNLKNTLNNLKGQPAKKANLIASKDNAKIHKQTL

SEQ ID NO:5

GQTTNKDAINDAIKDANNAKGTDKYNNASDDTKSKFDDALKKAEDVKNDSNANQK
EVDDATKNLKNTLNNLKGQPAKKANLIASKDNAKIHKQTL

SEQ ID NO:6

VTRTINVVDPITGKISTSVQTAKFTREDKNSNAGYTDPVTGKTTMNPWTPAKQGLRA
VNVEQIKGYVAKVDGNVDAVVVTPDSANMVVTITYQANKPEGQNITVKKDTVPDP
ADGIKNKDDLPDGTKYTWKEVPDVNSVGEKTGIVTVTFPDGTSVDVKVTVYVDPVV
ESNRDTLSKEANTGNTNVAKAATVTSSKVESKKT

SEQ ID NO:7

VTRTINVVDPITGKISTSVQTAKFTREDKNSNAGYTDPVTGKTTMNPWTPAKQGLRA
VNVEQIKGYVAKVDGNVDAVVVTPDSANMVVTITYQANKPEGQNITVKKDTVPDP
ADGIKNKDDLPDGTKYTWKEVPDVNSVGEKTGIVTVTFPDGTSVDVKVTVYVDPVV
ESNRDTLSKEANTGNTNVAKAATVTSSKVESKKT**LPQTG**SKTEQVGILGLAIATVGS
LLGLGVN

SEQ ID NO:8

KKAEEVKNNSNATQKEVDDATNNLKQAQNDLDGQTTDKSKLDEAIKSADDTKSTD
KYNNASDDTKSKFDEALKKAEEVKNNSNATQKEVDDATKNLKQAQNDLDGQTTN
KDAINDAIKDANNAKGTDKYNNASDDTKSKFDDALKKAEDVKNDSNANQKEVDD
ATKNLKNTLNNLKGQPAKKANLIASKDNAKIHKQTL**LPQTG**TETNPLTAIGIGLMAL
GAGIFA